

明 細 書

対訳判断装置、方法及びプログラム

技術分野

- [0001] 本発明は対訳判断装置、方法及びプログラムに係り、特に、原言語の原文中の少なくとも翻訳対象語句の対訳を判断する対訳判断装置、該対訳判断装置に適用可能な対訳判断方法、及びコンピュータを前記対訳判断装置として機能させるためのプログラムに関する。

背景技術

- [0002] コンピュータを利用して、或る自然言語(原言語)で記述された文章(原文)を、他の自然言語(目的言語)で記述された文章(翻訳文)に翻訳する、所謂機械翻訳の実現はかなり以前より期待されており、機械翻訳に関する様々な改良技術も提案されている。
- [0003] 例えば特許文献1には、HD装置に用意した英和連語辞書に、連語に代表される一まとまりの表現形態を格納しておき、統語解析処理において、等位接続詞によって結合された単語から構成される表現形態を英文テキスト中で検索し、検索した表現形態が英和連語辞書に格納されている場合、又は検索した表現形態を構成する単語の接頭辞又は接尾辞が同一である場合に、検索した表現形態を一の形態素として認識し、分離することなく構文の解析を行う技術が開示されている。
- [0004] また、特許文献2には、多義語毎に語義と特徴とを対応させた特徴テーブルを予め記憶しておき、入力された第1言語の原文について特徴レコードを生成し、生成した特徴レコードと特徴テーブルとを比較し原文が有する特徴に基づいて多義語の語義を選択・出力する技術が開示されている。

特許文献1:特開平11-328178号公報

特許文献2:特開平6-314294号公報

発明の開示

発明が解決しようとする課題

- [0005] 原言語で記述された原文を目的言語で記述された翻訳文に翻訳する作業は、通

常、文は単語と文法によって成り立っているという認識に基づき、原言語で記述された原文中の単語を目的言語の単語に変換し、変換した単語を目的言語の文法ルールに従って並び替える、という翻訳プロセスを経て行われる。この翻訳プロセスは、コンピュータを利用して翻訳を行う機械翻訳にも採用されており、対訳を単語単位で辞書に登録しておき、原文から順に単語を取り出し、取り出した単語の対訳を検索することを繰り返すことで、原文を単語単位で対訳に置き換えると共に、原文中の個々の単語の品詞を判断して構文を解析し、構文の解析結果に基づき目的言語の文法ルールに従って単語単位での対訳を並べ替えることで翻訳文(対訳文)を得る方式が一般的となっている。

[0006] また、機械翻訳において、原文中に定型的な連語が存在している場合にも、特許文献1に記載の技術を適用すれば、上記の連語の適切な対訳が得られる可能性があり、原文中に多義語が存在している場合にも、特許文献2に記載の技術を適用すれば、上記の多義語について特定の対訳が得られる可能性がある(但し、特許文献2に記載の技術では、多義語を抽出し、抽出した多義語の語義及び特徴を解析して特徴テーブルを予め作成しておく、という煩雑な作業が必要となる)。しかしながら、機械翻訳では、得られる翻訳文に文法的な誤りがなく、単語単位での対訳に誤りがない場合にも、翻訳文が目的言語の文として不自然な文となってしまうことが多々生じており、既存の機械翻訳の技術では、例え特許文献1や特許文献2に記載の技術を適用したとしても、実用に耐えうる翻訳精度が得られないのが実情である。

[0007] これは、機械翻訳で採用している翻訳プロセスが、人が母語で話したり文を書いたりするときに文を作成するプロセスと著しく相違していることが原因と推察される。すなわち、人が母語で話したり文を書いたりするときには、単語を思い浮かべ文法ルールに当てはめて文を作成する、という翻訳プロセスのようなプロセスは経ておらず、実際には、文脈の前後との関係や語句自体に付随する背景知識も踏まえた上で、個々の人の記憶に蓄積されている膨大な数のチャンク(意味のかたまり:単語・連語・定型表現・コロケーション・構文・文章)の中から、思い浮かんだ(状況に応じて選択した)適切な単語・連語・文を繋げることで文を作っている。

[0008] このため、人が母語で話したり書いたりする時と同じように、前後の文脈との関係、

語句自体に付随する背景知識、人が母語で話したり書いたりするための文書を作る時に思い浮かべる単位(すなわちチャンク)を踏まえた上で、原文の対訳を求めるようにすれば、目的言語の文として自然な翻訳文が得られることになるが、文脈の判断や語句自体に付随する背景知識の判断は容易でない上に、チャンクもその区切りが曖昧であり、機械翻訳において、チャンク単位で原文の対訳を求めて翻訳文を得ることは実現できていないのが実情であった。

[0009] 本発明は上記事実を考慮して成されたもので、原言語の原文から目的言語の文として自然な対訳文が得られる可能性の高い、原文中の翻訳対象語句の適切な対訳を得ることができる対訳判断装置、対訳判断方法及びプログラムを得ることが目的である。

課題を解決するための手段

[0010] 上記目的を達成するために請求項1記載の発明に係る対訳判断装置は、複数の単語から成る原言語の自然文を目的言語の対訳文と対応付けて複数記憶する記憶手段と、前記記憶手段に記憶されている原言語の複数の自然文の中から、原言語の原文中の翻訳対象語句が含まれている自然文を検索する検索手段と、前記検索手段による検索によって抽出された自然文と前記原文との一致度を求め、求めた一致度に基づいて選択した自然文の対訳文における少なくとも前記翻訳対象語句の対訳を、前記原文中の少なくとも前記翻訳対象語句の対訳と判断する第1の対訳判断手段と、を含んで構成されている。

[0011] 請求項1記載の発明では、複数の単語から成る原言語の自然文を目的言語の対訳文と対応付けて記憶手段に複数記憶している。なお、本発明に係る自然文は、従来の機械翻訳における辞書のように、単語単位での分割や多義語の抽出等の編集・加工を行っていない文、文節、連語、定型表現、コロケーションの少なくとも何れかであり、この自然文を目的言語の対訳文と対応付けて記憶することで、例えば単語単位で辞書に登録する場合のように、多義語に多数の対訳を対応付ける必要が無くなる(多義語に多数の対訳を対応付けた場合、対訳を選択する必要が生じ、選択誤りが生ずる可能性がある)と共に、自然文に対応する対訳文を目的言語の文として自然な文にすることができる。なお、本発明において、自然文以外に単語とその対訳も

記憶手段に記憶させても構わない。

[0012] また、請求項1記載の発明では、記憶手段に記憶されている原言語の複数の自然文の中から、原言語の原文(文、文節、連語、定型表現、コロケーションの何れでもよい)中の翻訳対象語句が含まれている自然文が検索手段によって検索される。翻訳対象語句は原文のうち特に翻訳すべき語句であり、単語であってもよいし、複数の単語で構成されていてもよい。例えば本発明に係る対訳判断装置を人間(翻訳者)が翻訳する際の電子辞書として用いる場合、翻訳対象語句は翻訳者によって指定される。また、例えば本発明に係る対訳判断装置を機械翻訳装置又は自動通訳装置の一部として用いる場合、翻訳対象語句は、本発明に係る対訳判断装置によって判断された対訳を利用して機械翻訳を行う機械翻訳装置又は自動通訳装置によって指定される。また、翻訳対象語句を含む原文についても、利用者(翻訳者等)に指定させるようにしてもよいし、自動的に判断する(例えば翻訳対象語句を含む文又は文節を自動的に原文と判断する等)ようにしてもよい。検索手段は、翻訳対象語句が含まれている自然文を検索するので、この検索により、対応する対訳文の中に翻訳対象語句の対訳が含まれている自然文が抽出されることになる。

[0013] また、請求項1記載の発明に係る第1の対訳判断手段は、検索手段による検索によって抽出された自然文と原文との一致度を求め、求めた一致度に基づいて選択した自然文の対訳文における少なくとも翻訳対象語句の対訳を、原文中の少なくとも翻訳対象語句の対訳と判断する。これにより、検索手段による検索によって抽出された自然文のうち、原文に近い自然文(例えば原文中に存在する多義語を同じ意味で用いている自然文等)が、高い確率で一致度の高い自然文として選択され、選択された自然文に対応する自然な対訳文が得られ、該対訳文における、少なくとも翻訳対象語句の対訳が、原文中の翻訳対象語句の対訳と判断されることになる(原文との一致度等に応じて、対訳文における翻訳対象語句以外の語句の対訳も、原文中の該語句の対訳と判断してもよいことは言うまでもない)。

[0014] このように、請求項1記載の発明は、原文上のチャンクの区切りを認識することが技術的に非常に困難であることに鑑み、原言語の自然文を目的言語の対訳文と対応付けて複数記憶しておき、記憶している自然文の中から原言語の原文との一致度の

高い自然文を選択し、選択した自然文に対応する自然な対訳文における対訳を、原文中の少なくとも翻訳対象語句の対訳と判断するので、原文上のチャンクの区切りを認識することなく、結果として、おおよそチャンク単位で原文の対訳を行ったに等しい自然な対訳を得ることができ、原文中の翻訳対象語句の適切な対訳(原文から目的言語の文として自然な対訳文が得られる可能性の高い対訳)を得ることができる。

[0015] また、請求項1記載の発明は、記憶手段に自然文と対訳文を記憶することで実現できるので、少なくとも記憶手段へ自然文及び対訳文を記憶させるにあたり、自然文を単語単位で分割したり、多義語を抽出したり、多義語について考え得る全ての対訳を列挙して各々対応付ける等の煩雑な編集・加工を行う必要がなくなる。また、請求項1記載の発明では、検索手段による検索によって抽出された自然文と原文との一致度を求め、求めた一致度に基づいて自然文を選択することで適切な対訳を得ることができるので、従来の機械翻訳における品詞判断や構文解析等の複雑な処理を行う必要がなくなり、処理を簡略化することができる。

[0016] なお、請求項1記載の発明では、記憶手段に自然文と対訳文を記憶するので、原文と完全一致する自然文が記憶手段に記憶されている可能性もある。これを考慮すると、例えば請求項2に記載したように、検索手段は、記憶手段に記憶されている原言語の複数の自然文の中から原文と完全一致している自然文も検索し、第1の対訳判断手段は、検索手段による検索によって原文と完全一致している自然文が抽出された場合に、当該完全一致している自然文の対訳文を原文の対訳文と判断することが好ましい。これにより、原文と完全一致する自然文が記憶手段に記憶されていた場合に、原文の対訳文を得ることができる。

[0017] また、請求項1記載の発明において、検索によって抽出された自然文と原文との一致度は、例えば以下のようにして求めることができる。すなわち、請求項3記載の発明は、請求項1記載の発明において、第1の対訳判断手段は、検索手段による検索によって抽出された自然文と原文との一致単語数を計数し、計数した一致単語数が多くなるに従って原文との一致度が高くなるように、原文との一致度を評価することを特徴としている。一致単語数は自然文と原文との一致度を表す重要な指標であり、一致単語数が多くなるに従って一致度が高くなるように一致度を評価することで、該一

致度に基づいて原文に近い自然文を精度良く選択することができる。

[0018] また、請求項3記載の発明において、一致度としては、例えば計数した一致単語数をそのまま用いることも可能であるが、例えば請求項4に記載したように、計数した一致単語数を、翻訳対象語句を構成する単語数で除した値を求め、求めた値を一致度として用いることが好ましい。これにより、本発明に係る一致度が、一致単語数を、翻訳対象語句を構成する単語数を基準として正規化した値になるので、この一致度を用いることで、翻訳対象語句を構成する単語数の多少に拘わらず、原文に近い自然文をより精度良く選択することができる。なお、一致単語数の計数にあたり、一致単語が翻訳対象語句を構成する単語かそれ以外の単語かを判別しておき、翻訳対象語句を構成する単語の一致単語数とそれ以外の単語の一致単語数に異なる重み(翻訳対象語句を構成する単語の方が重みが大きくなるように設定した重み)を乗じて加算した値(一致単語数評価値)を一致単語数として用いるようにしてもよい。

[0019] また、請求項3記載の発明において、第1の対訳判断手段は、請求項5に記載したように、検索手段による検索によって抽出された自然文と原文との不一致単語数も計数し、計数した不一致単語数が少なくなるに従って原文との一致度が高くなるように、原文との一致度を評価するようにしてもよい。不一致単語数も一致単語数と並んで自然文と原文との一致度を表す重要な指標であり、請求項3に記載の一致単語数に加えて上記の不一致単語も用い、不一致単語数が少なくなるに従って一致度が高くなるように一致度を評価することで、自然文と原文との一致度評価の正確性を更に向上させることができる。

[0020] ところで、本発明では記憶手段に自然文を記憶するので、請求項3記載の発明において、例えば英文における“a”, “the”, “to”, “in”等のように、原言語の自然文中に頻出する単語を一致単語と判断してしまうと、これらの頻出単語が多く含まれている原文については、頻出単語の影響により、実際には原文に近くない自然文が一致度の高い自然文として誤選択される可能性がある。これを考慮すると、第1の対訳判断手段は、例えば請求項6に記載したように、一致単語数の計数に際し、予め定められた頻出単語を計数対象から除外することが好ましい。これにより、頻出単語が一致単語数に及ぼす影響を排除することができ、一致単語数を、自然文と原文との一致度をよ

り正確に反映する指標として用いることができる。

[0021] また、請求項3又は請求項5記載の発明において、第1の対訳判断手段は、例えば請求項7に記載したように、一致単語数又は不一致単語数の計数に際し、単複又は時制の相違に起因して語尾が相違している単語を一致単語とみなして計数することが好ましい。単複又は時制の相違に起因して語尾が相違している単語は、例えば該単語を別途テーブルに登録しておき、語尾のみが不一致の単語については前記テーブルに登録されているか否かを判断する等によって認識することができる。これにより、本来は一致単語とみなすべき、単複又は時制の相違に起因して語尾が相違している単語が、一致単語数や不一致単語数に及ぼす影響を排除することができ、一致単語数や不一致単語数を、自然文と原文との一致度をより正確に反映する指標として用いることができる。なお、英文等における単語の大文字と小文字の相違等も一致単語とみなすことが好ましい。

[0022] また、請求項3記載の発明において、第1の対訳判断手段は、例えば請求項8に記載したように、一致単語数の計数に際し、複数回出現した一致単語を重複計数しないことが好ましい。これにより、複数回出現した一致単語が一致単語数に及ぼす影響を排除することができ、一致単語数を、自然文と原文との一致度をより正確に反映する指標として用いることができる。また、上記のように重複計数しないことに代えて、複数回出現した一致単語を予め定めた n ($n \geq 2$) 回以上計数しないようにしてもよい。

[0023] 更に、請求項3又は請求項5記載の発明において、第1の対訳判断手段は、例えば請求項9に記載したように、検索手段による検索によって抽出された自然文と原文との単語の並び順の類似度を評価し、単語の並び順の類似度が高くなるに従って原文との一致度が高くなるように、原文との一致度を評価することが好ましい。自然文では単語が同一であっても並び順によって意味(対訳)が相違する場合があるが、上記のように単語の並び順の類似度を評価し、単語の並び順の類似度が高くなるに従って一致度が高くなるように一致度を評価することで、該一致度に基づいて原文に近い自然文を精度良く選択することができる。

[0024] また、請求項3又は請求項5記載の発明において、第1の対訳判断手段は、例えば請求項10に記載したように、検索手段による検索によって抽出された自然文におい

て、原文との一致単語の間に存在する不一致単語数を計数し、計数した一致単語の間に存在する不一致単語数が少なくなるに従って原文との一致度が高くなるように、原文との一致度を評価することが好ましい。これにより、一致度の正確性が向上し、該一致度に基づいて原文に近い自然文を精度良く選択することができる。

[0025] なお、記憶手段にどのような自然文(及び対訳文)が記憶されているかによっても相違するが、請求項1記載の発明においても、例えば一致度が同一の自然文が複数抽出された等、一致度のみでは原文に近い適切な自然文を特定(選択)することが困難な場合も生じ得る。これを考慮すると、請求項1記載の発明において、例えば請求項11に記載したように、検索手段による検索によって抽出された自然文に基づいて、翻訳対象語句と原言語の同一の文中に出現する頻度の高い頻出語句を認識する認識手段を更に設け、第1の対訳判断手段は、検索手段による検索によって抽出された自然文のうち、認識手段によって認識されかつ原文中に存在する特定の頻出語句と、翻訳対象語句が各々含まれている自然文の対訳文を参照することで、翻訳対象語句及び特定の頻出語句が各々含まれる自然文の対訳文における翻訳対象語句についての高頻度の対訳を認識し、認識した高頻度の対訳を、原文中の翻訳対象語句の対訳と判断することが好ましい。

[0026] 翻訳対象語句と原言語の同一の文中に出現する頻度(コリレーションという)の高い特定の頻出語句が原文中に存在している場合、翻訳対象語句の対訳は、翻訳対象語句及び特定の頻出語句が各々含まれている自然文の対訳文における翻訳対象語句の対訳に一致している可能性が高い。但し、翻訳対象語句及び特定の頻出語句が各々含まれている自然文の対訳文における翻訳対象語句の適切な対訳は同じである可能性が高いものの、記憶手段に記憶されている上記の自然文の中に、翻訳対象語句の適切な対訳が相違している自然文が混在している可能性もある。上記に基づき請求項11記載の発明では、翻訳対象語句とのコリレーションの高い頻出語句を認識し、認識した頻出語句のうち原文中に存在する特定の頻出語句と、翻訳対象語句が各々含まれている自然文の対訳文を参照することで、翻訳対象語句及び特定の頻出語句が各々含まれる自然文の対訳文における翻訳対象語句についての高頻度の対訳を認識し、認識した高頻度の対訳を、原文中の翻訳対象語句の対訳と判

断するので、一致度のみでは適切な自然文を特定(選択)することが困難な場合にも、頻出語句に基づいて、原文中の翻訳対象語句の適切な対訳(原文から目的言語の文として自然な対訳文が得られる可能性の高い対訳)を得ることができる。

[0027] 訳文選択の際の前後の文脈の判断は、状況に応じて無数に存在し、事前に特定することは困難であるが、請求項11記載の発明(及び後述する請求項14記載の発明)では、同じ文章の中で動じに出現する確率の高い語句に着目し、その語句と翻訳対象語句が同時に出現する自然文の対訳文を参照することで、翻訳対象語句についての対訳を判断するので、原文の前後の文脈を認識することなく、結果として多義語についても、おおよそ前後の文脈を踏まえた上での適切な対訳を得ることができる。

[0028] なお、上記の頻出語句の認識は、例えば原言語の同一の文中に出現する頻度の高い語句同士をテーブルに登録しておき、該テーブルを参照することで行うことも可能であるが、請求項11記載の発明では、検索手段による検索によって抽出された自然文に基づいて頻出語句を認識しているので、上記のテーブル作成の手間が省けると共に、テーブルを記憶するために必要となる記憶容量を節減できる、という効果も得られる。

[0029] また、請求項1記載の発明において、例えば請求項12に記載したように、原文中に存在しかつ検索手段による検索によって抽出された自然文中に含まれていない着目語句と代替可能な代替語句を判断する判断手段を更に設け、第1の対訳判断手段は、検索手段による検索によって抽出された自然文のうち、判断手段によって判断された代替語句及び翻訳対象語句が各々含まれている自然文の対訳文における少なくとも翻訳対象語句の対訳を、原文中の少なくとも翻訳対象語句の対訳と判断することが好ましい。

[0030] 請求項1記載の発明において、検索手段による検索により、原文中の特定の語句が別の語句に入れ替わった自然文が抽出されることがあるが、このとき原文の意味と自然文の意味が類似であれば、特定の語句と別の語句は代替可能な関係にあると考えられる。そして、検索手段による検索によって、例えば原文との一致度が同一の自然文が複数抽出された場合にも、その中に原文に対して特定の語句が代替可能

な関係にある別の語句(代替語句)に入れ替わっている自然文が存在している場合には、この自然文は原文と意味が類似である可能性が高いので、この自然文を選択することが望ましい。

[0031] これに対して請求項12記載の発明では、原文中に存在しかつ検索手段による検索によって抽出された自然文中に含まれていない着目語句と代替可能な代替語句を判断手段によって判断し、第1の対訳判断手段は、検索手段による検索によって抽出された自然文のうち、判断手段によって判断された代替語句及び翻訳対象語句が各々含まれている自然文の対訳文における少なくとも翻訳対象語句の対訳を、原文中の少なくとも翻訳対象語句の対訳と判断するので、一致度のみでは適切な自然文を特定(選択)することが困難な場合にも、代替語句の有無に基づいて、原文中の翻訳対象語句の適切な対訳(原文から目的言語の文として自然な対訳文が得られる可能性の高い対訳)を得ることができる。

[0032] なお、請求項12記載の発明に係る判断手段による代替語句の判断は、例えば代替可能な関係にある語句同士をテーブルに登録しておき、このテーブルを参照することで行うようにしてもよいが、例えば請求項13に記載したように、着目語句を含む自然文を記憶手段に記憶されている複数の自然文の中から検索し、該検索によって抽出された自然文と同一の構文の自然文を記憶手段に記憶されている自然文の中から検索し、該検索によって抽出された自然文中で着目語句と置き換わっている語句を代替語句と判断するようにしてもよい。この場合も、上記のテーブル作成の手間が省けると共に、テーブルを記憶するために必要となる記憶容量を節減することができる。

[0033] 語句自体に付随する背景知識についても、語句によって様々な性質のものが無数に存在するので、事前にそれらを全て洗い出し、カテゴリーに応じて分類することは困難である。請求項12, 13記載の発明では、同じ構文の中で置き換え可能な語句を同じカテゴリーの語句と判断するので、あらゆる語句の背景知識を認識しカテゴリーに応じて分類することなく、おおよそ語句自体に付随する背景知識を踏まえた上での適切な対訳を得ることができる。

[0034] 請求項14記載の発明に係る対訳判断装置は、複数の単語から成る原言語の自然

文を目的言語の対訳文と対応付けて複数記憶する記憶手段と、前記記憶手段に記憶されている原言語の複数の自然文の中から、原言語の原文中の翻訳対象語句が含まれている自然文を検索する検索手段と、前記検索手段による検索によって抽出された自然文に基づいて、前記翻訳対象語句と原言語の同一の文中に出現する頻度の高い頻出語句を認識する認識手段と、前記検索手段による検索によって抽出された自然文のうち、前記認識手段によって認識されかつ前記原文中に存在する特定の頻出語句と、前記翻訳対象語句が各々含まれている自然文の対訳文を参照することで、前記翻訳対象語句及び前記特定の頻出語句が各々含まれる自然文の対訳文における翻訳対象語句についての高頻度の対訳を認識し、認識した高頻度の対訳を、前記原文中の少なくとも前記翻訳対象語句の対訳と判断する第2の対訳判断手段と、を含んで構成されている。

[0035] 請求項14記載の発明では、請求項1記載の発明と同様の記憶手段及び検索手段が設けられており、認識手段は、検索手段による検索によって抽出された自然文に基づいて、翻訳対象語句と原言語の同一の文中に出現する頻度の高い頻出語句を認識し、第2の対訳判断手段は、検索手段による検索によって抽出された自然文のうち、認識手段によって認識されかつ原文中に存在する特定の頻出語句と、翻訳対象語句が各々含まれている自然文の対訳文を参照することで、翻訳対象語句及び特定の頻出語句が各々含まれる自然文の対訳文における翻訳対象語句についての高頻度の対訳を認識し、認識した高頻度の対訳を、原文中の翻訳対象語句の対訳と判断する。これにより、請求項11記載の発明と同様に、記憶手段に記憶され、かつ翻訳対象語句及び特定の頻出語句が各々含まれている自然文の中に、翻訳対象語句の適切な対訳が相違している自然文が混在していた場合にも、この影響を受けることなく、原文中の翻訳対象語句の適切な対訳(原文から目的言語の文として自然な対訳文が得られる可能性の高い対訳)を得ることができる。

[0036] また、請求項1又は請求項14記載の発明において、翻訳対象語句は複数の単語から構成されていてもよいが、翻訳対象語句を構成する単語数が多くなってくると、検索手段が検索を行っても翻訳対象語句が全て含まれている自然文が抽出されない事態が生ずる可能性もある。これを考慮すると、例えば請求項15に記載したように

、翻訳対象語句が複数の単語から成る場合、検索手段は、記憶手段に記憶されている原言語の複数の自然文の中から、翻訳対象語句を構成する複数の単語のうちの少なくとも1つが含まれている自然文を検索する(すなわち、翻訳対象語句が全て含まれている自然文及び翻訳対象語句の一部が含まれている自然文を各々検索する)ことが好ましい。これにより、翻訳対象語句として多数の単語から成る語句が指定されたことで、翻訳対象語句が全て含まれている自然文が記憶手段に記憶されていない場合にも、検索手段による検索により、少なくとも翻訳対象語句の対訳を推定可能な自然文(対訳文)を得ることができる。

[0037] 請求項16記載の発明に係る対訳判断方法は、目的言語の対訳文と対応付けて記憶手段に複数記憶されている、各々複数の単語から成る原言語の自然文の中から、原言語の原文中の翻訳対象語句が含まれている自然文を検索する第1のステップ、及び、前記第1のステップにおける検索によって抽出された自然文と前記原文との一致度を求め、求めた一致度に基づいて選択した自然文の対訳文における少なくとも前記翻訳対象語句の対訳を、原文中の前記翻訳対象語句の対訳と判断する第2のステップを含んでいるので、請求項1記載の発明と同様に、原文中の翻訳対象語句の適切な対訳(原文から目的言語の文として自然な対訳文が得られる可能性の高い対訳)を得ることができる。

[0038] 請求項17記載の発明に係る対訳判断方法は、目的言語の対訳文と対応付けて記憶手段に複数記憶されている、各々複数の単語から成る原言語の自然文の中から、原言語の原文中の翻訳対象語句が含まれている自然文を検索する第1のステップ、前記第1のステップにおける検索によって抽出された自然文に基づいて、前記翻訳対象語句と原言語の同一の文中に出現する頻度の高い頻出語句を認識する第2のステップ、前記第1のステップにおける検索によって抽出された自然文のうち、前記第2のステップで認識されかつ前記原文中に存在する特定の頻出語句と、前記翻訳対象語句が各々含まれている自然文の対訳文を参照することで、前記翻訳対象語句及び前記特定の頻出語句が各々含まれる自然文の対訳文における翻訳対象語句についての高頻度の対訳を認識し、認識した高頻度の対訳を、原文中の前記翻訳対象語句の対訳と判断する第3のステップを含んでいるので、請求項14記載の発明

と同様に、原文中の翻訳対象語句の適切な対訳(原文から目的言語の文として自然な対訳文が得られる可能性の高い対訳)を得ることができる。

[0039] 請求項18記載の発明に係るプログラムは、複数の単語から成る原言語の自然文を目的言語の対訳文と対応付けて複数記憶する記憶手段と接続されたコンピュータを、前記記憶手段に記憶されている原言語の複数の自然文の中から、原言語の原文中の翻訳対象語句が含まれている自然文を検索する検索手段、及び、前記検索手段による検索によって抽出された自然文と前記原文との一致度を求め、求めた一致度に基づいて選択した自然文の対訳文における少なくとも前記翻訳対象語句の対訳を、前記原文中の少なくとも前記翻訳対象語句の対訳と判断する第1の対訳判断手段、として機能させる。

[0040] 請求項18記載の発明に係るプログラムは、複数の単語から成る原言語の自然文を目的言語の対訳文と対応付けて複数記憶する記憶手段と接続されたコンピュータ(記憶手段を内蔵しているコンピュータであってもよいし、記憶手段と接続された別のコンピュータと通信回線を介して接続されているコンピュータであってもよい)を、上記の検索手段及び第1の対訳判断手段として機能させるためのプログラムであるので、上記コンピュータが請求項18記載の発明に係るプログラムを実行することにより、コンピュータが請求項1に記載の対訳判断装置として機能することになり、請求項1記載の発明と同様に、原文中の翻訳対象語句の適切な対訳(原文から目的言語の文として自然な対訳文が得られる可能性の高い対訳)を得ることができる。

[0041] 請求項19記載の発明に係るプログラムは、複数の単語から成る原言語の自然文を目的言語の対訳文と対応付けて複数記憶する記憶手段と接続されたコンピュータを、前記記憶手段に記憶されている原言語の複数の自然文の中から、原言語の原文中の翻訳対象語句が含まれている自然文を検索する検索手段、前記検索手段による検索によって抽出された自然文に基づいて、前記翻訳対象語句と原言語の同一の文中に出現する頻度の高い頻出語句を認識する認識手段、及び、前記検索手段による検索によって抽出された自然文のうち、前記認識手段によって認識されかつ前記原文中に存在する特定の頻出語句と、前記翻訳対象語句が各々含まれている自然文の対訳文を参照することで、前記翻訳対象語句及び前記特定の頻出語句が

各々含まれる自然文の対訳文における翻訳対象語句についての高頻度の対訳を認識し、認識した高頻度の対訳を、前記原文中の前記翻訳対象語句の対訳と判断する第2の対訳判断手段として機能させる。

- [0042] 請求項19記載の発明に係るプログラムは、複数の単語から成る原言語の自然文を目的言語の対訳文と対応付けて複数記憶する記憶手段と接続されたコンピュータ(記憶手段を内蔵しているコンピュータであってもよいし、記憶手段と接続された別のコンピュータと通信回線を介して接続されているコンピュータであってもよい)を、上記の検索手段、認識手段及び第2の対訳判断手段として機能させるためのプログラムであるので、上記コンピュータが請求項19記載の発明に係るプログラムを実行することにより、コンピュータが請求項14に記載の対訳判断装置として機能することになり、請求項14記載の発明と同様に、原文中の翻訳対象語句の適切な対訳(原文から目的言語の文として自然な対訳文が得られる可能性の高い対訳)を得ることができる。

発明の効果

- [0043] 以上説明したように本発明は、記憶手段に記憶されている原言語の複数の自然文の中から、原言語の原文中の翻訳対象語句が含まれている自然文を検索し、該検索によって抽出された自然文と原文との一致度を求め、求めた一致度に基づいて選択した自然文の対訳文における少なくとも翻訳対象語句の対訳を、原文中の少なくとも翻訳対象語句の対訳と判断するので、原言語の原文から目的言語の文として自然な対訳文が得られる可能性の高い、原文中の翻訳対象語句の適切な対訳を得ることができる、という優れた効果を有する。
- [0044] また本発明は、記憶手段に記憶されている原言語の複数の自然文の中から、原言語の原文中の翻訳対象語句が含まれている自然文を検索し、該検索によって抽出された自然文に基づいて、翻訳対象語句と原言語の同一の文中に出現する頻度の高い頻出語句を認識し、原文中に存在する特定の頻出語句と翻訳対象語句が各々含まれている自然文の対訳文を参照することで、翻訳対象語句及び特定の頻出語句が各々含まれる自然文の対訳文における翻訳対象語句についての高頻度の対訳を認識し、認識した高頻度の対訳を、原文中の前記翻訳対象語句の対訳と判断するので、原言語の原文から目的言語の文として自然な対訳文が得られる可能性の高い、

原文中の翻訳対象語句の適切な対訳を得ることができる、という優れた効果を有する。

図面の簡単な説明

[0045] [図1]本実施形態に係るPCの概略構成を示すブロック図である。

[図2]対訳判断処理の内容を示すフローチャートである。

[図3]対訳DBに登録されている“operation”を含む自然文と対訳文の一例を示す図表である。

[図4]ネットワークを介してクライアントPCとサーバ・コンピュータが接続されたコンピュータ・システムに本発明を適用した態様を説明するためのブロック図である。

[図5]本発明を適用した機械翻訳装置の概略構成の一例を示すブロック図である。

発明を実施するための最良の形態

[0046] 以下、図面を参照して本発明の実施形態の一例を詳細に説明する。なお、以下では、まずユーザから翻訳対象語句が指定されると、指定された翻訳対象語句の対訳をユーザに提示する電子辞書装置に本発明を適用した態様について説明する。

[0047] 図1には、上記の電子辞書装置として機能することが可能なパーソナル・コンピュータ(PC)10が示されている。PC10はCPU10A、ROM10B、RAM10C及び入出力ポート10Dを備えており、これらがデータバス、アドレスバス、制御バス等から成るバス10Eを介して互いに接続されて構成されている。また入出力ポート10Dには、各種の入出力機器として、CRTやLCD等から成るディスプレイ12、ユーザがデータ等を入力するためキーボード14、マウス16、ハードディスクドライブ(HDD)18、CD-ROM24からのデータの読み出しを行うCD-ROMドライブ20、及び紙原稿等の読み取りが可能なスキャナ22が各々接続されている。

[0048] PC10には、PC10を電子辞書装置として機能させるための対訳判断プログラム(請求項18及び請求項19に記載のプログラムに相当)がHDD18にインストールされており、このHDD18には、前記対訳判断プログラムが対訳判断に用いるためのデータが記憶された対訳データベース(対訳DB)も記憶されている。対訳判断プログラムをPC10にインストール(移入)するには幾つかの方法があるが、例えば対訳判断プログラムをセットアッププログラムと共にCD-ROM24に記録しておき、該CD-ROM2

4をCD-ROMドライブ20にセットし、CPU10Aに対して前記セットアッププログラムの実行を指示すれば、CD-ROM24から対訳判断プログラムが順に読み出され、読み出された対訳判断プログラムがHDD18に順に書き込まれることで、対訳判断プログラムのインストールが行われる。

- [0049] また、対訳DBについても、例えば予めCD-ROM24に記録しておき、対訳判断プログラムのインストール時に同時にHDD18に書き込まれるようにセットアッププログラムを構成することで、HDD18に記憶させることができる。対訳DBには、複数の単語から成り原言語で記述された自然文(単語単位での分割や多義語の抽出等の編集・加工を行っていない文、文節、連語、定型表現、コロケーションの何れか、具体例は後述する)のテキストデータが、目的言語で記述された対訳文(この対訳文も単語単位での分割や多義語の抽出等の編集・加工を行っていない自然文である)のテキストデータと対応付けられて多数登録されており、この態様において、対訳DBを記憶するHDD18は本発明に係る記憶手段に対応している。
- [0050] なお、対訳DBは、CD-ROM24やDVD-ROM等の記録媒体に記録しておき、この記録媒体から直接データを読み出すことで利用することも可能であり、この態様では対訳DBが記録されている記録媒体が本発明に係る記憶手段として機能することになる。また、本実施形態に係る対訳DBには、原言語の単語と目的言語の対訳も対応付けられて登録されている。本実施形態に係る対訳DBは、例えば原言語の単語と目的言語の対訳が対応付けられた既存の辞書に、自然文とその対訳文を適宜追加していくことで作成することができる。
- [0051] 次に本実施形態の作用を説明する。本実施形態において、原言語で記述された原文の中に目的言語での対訳を知りたい語句(翻訳対象語句:単語でもよいし、原文中の連続する複数の単語から成る語句であってもよい)が存在していることを認識した場合、ユーザはPC10を介して原文中の翻訳対象語句の対訳の出力を指示する所定の操作を行う。なお、原文としては、PC10にテキストデータとして読み込まれているテキスト(例えばユーザがキーボード14を介して入力したテキスト、ワープロソフトによって作成されたテキスト、インターネットを介して閲覧中のウェブページ内のテキスト、文字原稿をスキャナによって読み取りOCR(Optical Character Recognition

: 光学的手法による文字認識) 処理を経て得られたテキスト等) を適用することができる。

[0052] また、上記の所定の操作としては、例えばディスプレイ12に原文が表示されている状態で、翻訳対象語句を選択することで翻訳対象語句を反転表示させた後に、翻訳対象語句を右クリックすることで表示されるコンテキストメニュー内の「対訳出力」に相当する項目を選択する等の操作を適用することができる。また、上記のようにコンテキストメニューを利用することに代えて、例えばツールバー内に表示されている所定のアイコンを選択する等の操作を行うことで、翻訳対象語句の対訳の出力を指示可能とすることも可能である。翻訳対象語句の対訳の出力を指示する上記のような操作が行われると、PC10のCPU10Aによって対訳判断プログラムが実行されることで、図2に示す対訳判断処理が行われる。なお、この対訳判断処理は請求項16及び請求項17に記載の対訳判断方法が適用された処理であり、この処理を行うことでPC10は電子辞書装置(請求項1及び請求項14に記載の対訳判断装置)として機能する。

[0053] 対訳判断処理では、まずステップ100において、指定された翻訳対象語句を含む単一の原文(処理対象の原文)のテキストデータを取り込むと共に、取り込んだ処理対象の原文中の翻訳対象語句を識別する情報を取り込む。なお、処理対象の原文は、翻訳対象語句を含む文であってもよいし、翻訳対象語句を含む文節、連語、定型表現、コロケーションの何れであってもよいが、文節、連語、定型表現、コロケーションの何れかを処理対象の原文とする場合、処理対象の原文としての文節、連語、定型表現、コロケーションは、ユーザによって指定させるようにしてもよいし、対訳判断処理で自動的に判断することも可能である。

[0054] 次のステップ102では、ステップ100で取り込んだ処理対象の原文のテキストデータをキーにして、対訳DBに登録されている自然文の中から処理対象の原文と完全一致している自然文を検索すると共に、翻訳対象語句のテキストデータをキーにして、対訳DBに登録されている自然文の中から翻訳対象語句が含まれている自然文(少なくとも翻訳対象語句が含まれていることで原文と部分一致している自然文)を検索する。また、翻訳対象語句が複数の単語から構成されている場合には、翻訳対象

語句を構成する複数の単語のうちの少なくとも1つが含まれている自然文も同時に検索される。なお、ステップ102は、請求項1（詳しくは請求項2及び請求項15）、請求項14, 18, 19に記載の検索手段に対応しており、請求項16, 17に記載の第1のステップにも対応している。次のステップ104以降は請求項1, 18に記載の第1の対訳判断手段に対応している。

[0055] ステップ104では、ステップ102の検索により処理対象の原文と完全一致している自然文が対訳DBから抽出されたか否か判定する。この判定が肯定された場合はステップ106へ移行し、処理対象の原文と完全一致している自然文と対応付けられて対訳DBに登録されている目的言語の対訳文を読み出してディスプレイ12に表示し、処理を終了する。なお、この対訳文の表示に際しては、読み出した対訳文上での翻訳対象語句の対訳が認識され、認識された翻訳対象語句の対訳が強調表示される。この場合、ユーザは指定した翻訳対象語句の適切な対訳（原文から目的言語の文として自然な対訳文が得られる可能性の高い対訳）を認識できると共に、指定した翻訳対象語句を含む処理対象の原文の適切な対訳文（目的言語の文として自然な対訳文）を認識することができる。なお、上記のステップ104, 106は請求項2に記載の第1の対訳判断手段に対応している。

[0056] （例1）上記処理について、実例を挙げて更に説明する。原言語が英語、目的言語が日本語であり、翻訳対象語句として、

“For your safety, don’t rush into the train.”

という処理対象の原文に含まれる“For your safety”が指定された場合、通常の翻訳プロセスでは、原言語と対訳を単語単位で登録しているので、上記の翻訳対象語句を構成する単語“safety”に対しては対訳「安全」が選択される。しかし、本実施形態に係る対訳判断処理では、対訳DBに原言語の自然文と目的言語の対訳文を対応付けて登録しているので、上記の原文と完全一致している自然文が対訳DBに登録されている可能性がある。そして、完全一致している自然文が対訳DBに登録されていた場合、当該自然文と対応付けられて対訳DBに登録されている対訳文が、以下のように翻訳対象語句の対訳が強調表示されて出力される。

「危険ですから、駆け込み乗車はおやめ下さい。」

通常の翻訳プロセスでは、単語"safety"の対訳として「危険」を選択することは困難であり、「あなたの安全のため、その列車に駆け込まないでください。」等の不自然な対訳文しか得られないが、本発明によれば母語として自然な対訳文が得られる。

[0057] なお、ステップ106における対訳文の表示において、処理対象の原文と完全一致している自然文に加えて、ステップ102の検索によって抽出された処理対象の原文と部分一致している自然文も併せて表示するようにしてもよい。処理対象の原文と部分一致している自然文は検索によって通常複数抽出されるが、これらを表示する場合には、処理対象の原文の部分一致している個々の自然文に対して次に述べる一致度を各々演算し、各自然文の対訳文を一致度の降順にディスプレイ12に表示することが好ましい。

[0058] 一方、ステップ102の検索により処理対象の原文と完全一致している自然文が対訳DBから抽出されなかった場合には、ステップ104の判定が否定されてステップ108へ移行し、ステップ102の検索により対訳DBから抽出された各自然文(処理対象の原文と部分一致している各自然文)について、自然文の各単語のうち原文中の何れかの単語と一致する単語の数(一致単語数)を各々計数し、一致単語数の計数結果に基づいて原文との一致度を各々演算する。

[0059] なお、本実施形態では、対訳判断プログラムのインストール時に、原言語の自然文中に頻出する単語(例えば英語の自然文における"a","the","to","in"等)が登録された頻出単語テーブルもHDD18に記憶される。そして、ステップ108における一致単語数の計数は頻出単語テーブルを参照しながら行われ、頻出単語テーブルに登録されている単語は一致単語数の計数対象から除外される。これにより、頻出単語が一致単語数に及ぼす影響を排除することができる。なお、ステップ108における上記処理は請求項6に記載の第1の対訳判断手段に対応している。

[0060] また、本実施形態では、対訳判断プログラムのインストール時に、単複又は時制の相違により語尾が相違している単語が登録された語尾変化単語テーブルもHDD18に記憶される。そして、ステップ108における一致単語数の計数では、語尾のみが不一致の単語が出現した場合に、語尾変化単語テーブルが参照されることで、語尾の不一致が単複又は時制の相違によるものか否かが判断され、単複又は時制の相違

により語尾が相違している単語は一致単語とみなして計数される。なお、非定型の語尾変化のみ語尾変化単語テーブルに登録し、定型の語尾変化(例えば英単語における複数形の"s"や過去形の"ed"等)は自動的に一致単語と判断するようにしてもよい。これにより、単複又は時制の相違により語尾が相違している単語(本来は一致とみなすべき単語)が一致単語数に及ぼす影響も排除することができる。なお、ステップ108における上記処理は請求項7に記載の第1の対訳判断手段に対応している。

[0061] また、本実施形態では、ステップ108における一致単語数の計数に際し、原文中の何れかの単語と一致したことで一致単語数に一旦計上した単語は、自然文中に再度出現したとしても一致単語数に計上しないことにより、複数回出現した一致単語を重複計数しないようにしている。これにより、同一の一致単語が自然文中の複数箇所に存在していた場合にも、この一致単語が一致単語数に及ぼす影響を排除することができる。なお、ステップ108における上記処理は請求項8に記載の第1の対訳判断手段に対応している。

[0062] また、一致度の演算式は、一致単語数が多くなるに従って一致度が高くなるように定めることができ、例えば一致単語数を、翻訳対象語句を構成する単語数で正規化する演算式(一致度＝一致単語数÷翻訳対象語句を構成する単語数)を用いることができる。上記のステップ108は請求項3(詳しくは請求項4)に記載の第1の対訳判断手段に対応している。

[0063] 次のステップ110では、各自然式について演算した一致度を比較することで、一致度が最大の自然文が複数存在しているか否か判定する。判定が否定された場合にはステップ112へ移行し、一致度最大の自然文と対応付けられて対訳DBに登録されている目的言語の対訳文を読み出し、読み出した対訳文上での翻訳対象語句の対訳を認識し、認識された翻訳対象語句の対訳が強調表示されるように、読み出した対訳文をディスプレイ12に表示して処理を終了する。この場合、ユーザは指定した翻訳対象語句の適切な対訳(原文から目的言語の文として自然な対訳文が得られる可能性の高い対訳)を認識することができる。

[0064] なお、ステップ112では、ステップ102の検索によって抽出された自然文のうち、一致度が最大の自然文以外に、一致度の降順に複数の自然文(一致度の降順に一定

数の自然文、或いは一致度が所定値以上の全ての自然文)の対訳文も対訳DBから読み出し、ディスプレイ12に併せて一覧表示する。また、一致度が同一の自然文が複数存在していた場合には、各自然文における原文との不一致単語数を計数し、対訳文の一覧表示に際して、一致度が同一の自然文に対応する対訳文を、対応する自然文における不一致単語数の昇順(少ない順)に表示する。

[0065] (例2) 上記処理について実例を挙げて更に説明する。原言語が英語、目的言語が日本語であり、翻訳対象語句として、

"The system suddenly came into operation."

という処理対象の原文に含まれる"operation"が指定されると共に、翻訳対象語句"operation"を含む自然文として図3に示す自然文(1)～(8)が対訳文(1)～(8)と対応付けられて対訳DBに登録されていた場合、図3に示す自然文(3)以外は原文との一致単語数が「1」で一致度が100%、(自然文(2)の"operations"も前述のように一致単語として計数される)、自然文(3)は原文との一致単語数が「4」(単語"the"は頻出として一致単語数の計数対象から除外される)で一致度が400%となるので、自然文(3)と対応付けられて対訳DBに登録されている対訳文(3)が、以下のように翻訳対象語句の対訳が強調表示されて出力される。

「システムが稼働し始めた。」

(例3) 原言語が英語、目的言語が日本語であり、翻訳対象語句として、

"A small business is exempt from the operation of the new law."

という処理対象の原文に含まれる"operation"が指定されると共に、翻訳対象語句"operation"を含む自然文として図3に示す自然文(1)～(8)が対訳文(1)～(8)と共に対訳DBに登録されていた場合、自然文(2)以外は原文との一致単語数が「1」で一致度が100%、自然文(2)は原文との一致単語数が「4」(単語"the"は頻出単語として一致単語数の計数対象から除外される)で一致度が400%となるので、自然文(2)と対応付けられて対訳DBに登録されている対訳文(2)が、以下のように翻訳対象語句の対訳が強調表示されて出力される。

「新法の適用」

英文における"operation"は多義語であり、適切な対訳の選択が困難な単語であるが

、上述した例2及び例3からも明らかなように、対訳DBに自然文と対訳文を対応付けて登録しておき、一致度に基づいて自然文(対訳文)の選択を行うことで、多義語を原文と同じ意味で用いている適切な自然文(対訳文)を選択することができる。

[0066] (例4) 中国語の文は、中国語を熟知していないと意味の固まりが判別しにくく、他の言語の文への翻訳に際し、対訳を求めるべき単位を判断することが困難なことが多い。これに対し、本実施形態に係る対訳判断処理では、対訳DBに登録されている自然文のうち、指定された翻訳対象語句が含まれている自然文を検索し(翻訳対象語句が複数の単語から構成されている場合は、翻訳対象語句を構成する複数の単語のうちの少なくとも1つが含まれている自然文も同時に検索し)、原文全体との一致度が高い自然文の対訳文を(一致度最大の自然文以外の自然文の対訳文も)表示するので、表示された対訳文を参照することで、ユーザが、原文上での対訳を求めるべき単位と、その適切な対訳を同時に認識できる。

[0067] 例えば、原言語としての中国語の原文「発展中国家」を、目的言語としての日本語に翻訳する場合を例にすると、翻訳対象語句として「発」「発展」「発展中」「発展中国」「発展中国家」の何れが指定された場合にも、一致度に基づき以下の自然文の対訳文が以下の順序で一覧表示される(なお、以下では翻訳対象語句として「発展中国家」が指定された場合の一致度も併せて示すが、個々の一致度は翻訳対象語句として指定された語句に応じて変化することは言うまでもない)。

1. 「発展中国家」=「発展途上国」 (一致単語数「5」、一致度100%)
2. 「発展中国」=「中国を発展させる」 (一致単語数「4」、一致度80%)
3. 「発展中」=「発展中の」 (一致単語数「3」、一致度60%)
4. 「発展」=「発展する、発展させる」(一致単語数「2」、一致度40%)
5. 「発」=「発する、出す」 (一致単語数「1」、一致度20%)

これにより、ユーザは一覧表示された上記の対訳文を参照することで、原文上での対訳を求めるべき単位(この場合は「発展中国家」)と、その適切な対訳を同時に認識することができる。

[0068] なお、原文が「発」で翻訳対象語句も「発」の場合、上記各自然文は一致単語数「1」で一致度100%になるが、不一致単語数が各々相違しているため、原文が「発展中

国家」の場合とは異なる以下の順序で表示される。

- 1.「発」 (不一致単語数「0」)
- 2.「発展」 (不一致単語数「1」)
- 3.「発展中」 (不一致単語数「2」)
- 4.「発展中国」 (不一致単語数「3」)
- 5.「発展中国」 (不一致単語数「4」)

同様に、原文が「発展」で翻訳対象語句が「発展」の場合にも、一致度及び不一致単語数に基づき、原文が「発展中国」や「発」の場合とは異なる以下の順序で表示される。

- 1.「発展」 (一致単語数「2」、一致度100%、不一致単語数「0」)
- 2.「発展中」 (一致単語数「2」、一致度100%、不一致単語数「1」)
- 3.「発展中国」 (一致単語数「2」、一致度100%、不一致単語数「2」)
- 4.「発展中国」 (一致単語数「2」、一致度100%、不一致単語数「3」)
- 5.「発」 (一致単語数「1」、一致度50%、不一致単語数「0」)

このように、翻訳対象語句が同一であっても原文が相違していれば原文に応じた適切な順序で対訳文が一覧表示されるので、ユーザは対訳文の一覧表示結果を参照することで、原文に応じた適切な対訳を認識することができる。

[0069] また、一致度が最大の自然文が複数存在していた場合、ステップ110の判定が肯定されてステップ114へ移行し、ステップ114, 116において、翻訳対象語句と原言語の同一の文中に出現する頻度の高い単語(翻訳対象語句とのコリレーションの高い単語:請求項11, 14, 17, 19に記載の頻出語句に相当)に基づいて自然文を選択するコリレーション解析を行う。

[0070] すなわち、ステップ114では、ステップ102の検索により対訳DBから抽出された各自然文に対し、処理対象の原文中の翻訳対象語句以外の各単語(但し、頻出単語テーブルに登録されている頻出単語は除外)が含まれているか否かを検索することで、翻訳対象語句と原言語の同一の文中に出現する頻度の高い単語(翻訳対象語句とのコリレーションの高い単語)が処理対象の原文に含まれているか否かを探索する。この処理では、処理対象の原文中の翻訳対象語句以外の各単語のうち、例えば

各自然文中における出現回数が所定回(1回でも複数回でもよい)以上の単語を、翻訳対象語句とのコリレーションの高い単語と判断することができる。なお、上記のステップ114は請求項11及び請求項14, 19に記載の認識手段に対応しており、請求項17に記載の第2のステップにも対応している。

[0071] 次のステップ116では、ステップ114における探索により翻訳対象語句とのコリレーションの高い単語が見つかったか否か判定する。判定が肯定された場合はステップ118へ移行し、ステップ102の検索により対訳DBから抽出された各自然文のうち、翻訳対象語句及びステップ114で判断した翻訳対象語句とのコリレーションが高い単語(処理対象の原文中に存在している単語)が各々含まれている自然文について、その対訳文を参照し、上記の自然文における翻訳対象語句についての高頻度の対訳を判断する。なお、翻訳対象語句についての高頻度の対訳は、具体的には、翻訳対象語句及び翻訳対象語句とのコリレーションが高い単語が各々含まれている自然文として単一の自然文が抽出された場合には、該自然文における翻訳対象語句の対訳を高頻度の対訳と判断し、上記の自然文として複数の自然文が抽出された場合には、これらの自然文における翻訳対象語句の対訳のうち最も出現頻度の高い対訳を高頻度の対訳と判断することができる。

[0072] 上記のように、翻訳対象語句が含まれており、かつ翻訳対象語句とのコリレーションが高い単語として処理対象の原文と同一の単語が含まれている自然文が存在している場合、当該自然文は翻訳対象語句を処理対象の原文と同じ意味で用いている文である可能性が高いが、上記の自然文の中に、翻訳対象語句についての対訳が相違している自然文が混在している可能性もある。これに対してステップ118では、翻訳対象語句及び翻訳対象語句とのコリレーションが高い単語が各々含まれている自然文の対訳文を参照して、上記の自然文における翻訳対象語句についての高頻度の対訳を判断するので、処理対象の原文における翻訳対象語句の適切な対訳を得ることができる。

[0073] そしてステップ119では、翻訳対象語句及び翻訳対象語句とのコリレーションが高い単語が各々含まれている自然文のうち、翻訳対象語句がステップ118で認識した高頻度の対訳と対応付けられている自然文と対応付けられて対訳DBに登録されて

いる目的言語の対訳文を読み出し、読み出した対訳文上での翻訳対象語句及び翻訳対象語句とのコリレーションが高い単語の対訳が強調表示されるように、読み出した対訳文をディスプレイ12に表示して処理を終了する。この場合もユーザは指定した翻訳対象語句の適切な対訳(原文から目的言語の文として自然な対訳文が得られる可能性の高い対訳)を認識することができる。なお、上記のステップ116〜ステップ119は、請求項11に記載の第1の対訳判断手段、請求項14, 19に記載の第2の対訳判断手段及び請求項17に記載の第3のステップに各々対応している。

[0074] なお、ステップ119においても、ステップ102の検索によって抽出された自然文のうち、翻訳対象語句及び翻訳対象語句とのコリレーションが高い単語が各々含まれており、翻訳対象語句が認識した高頻度の対訳と対応付けられている自然文以外に、一致度の降順に複数の自然文(一致度の降順に一定数の自然文、或いは一致度が所定値以上の全ての自然文)の対訳文も対訳DBから読み出し、一致度及び不一致単語数に応じた順序でディスプレイ12に併せて一覧表示する。

[0075] (例5) 上記処理について実例を挙げて更に説明する。原言語が英語、目的言語が日本語であり、翻訳対象語句として、

"I had an operation to remove a rectal cancer."

という処理対象の原文に含まれる"operation"が指定されると共に、翻訳対象語句"operation"を含む自然文として図3に示す自然文(1)〜(8)及び他の自然文が各々対訳文と共に対訳DBに登録されていた場合、自然文(6),(8)は何れも原文との一致単語数が「2」であるので、一致度のみから最適な自然文(対訳文)を選択することは困難である。ここで、処理対象の原文中の翻訳対象語句"operation"以外の各単語(例えば"remove","rectal","cancer"等)をキーにして翻訳対象語句"operation"を含む自然文(1)〜(8)を検索すると、自然文(6),(8)に"cancer"が含まれていることから、処理対象の原文における翻訳対象語句"operation"とのコリレーションの高い単語として"cancer"が抽出される。

[0076] ここで、翻訳対象語句"operation"及びコリレーションが高い単語"cancer"が含まれている自然文は翻訳対象語句"operation"を処理対象の原文と同じ意味で用いている文である可能性が高いが、対訳DBに登録されている上記の条件を満たす自然文

の中に、翻訳対象語句についての対訳が処理対象の原文と相違している自然文が混在している可能性もある。本例では自然文(8)がそれに該当する。自然文(8)は翻訳対象語句“operation”及びコリレーションが高い単語“cancer”が含まれているものの、翻訳対象語句“operation”の対訳が「操作」であるので、処理対象の原文と相違している。

- [0077] このような自然文の影響による翻訳対象語句の対訳の誤判断を防止するために、翻訳対象語句“operation”及びコリレーションが高い単語“cancer”が含まれている自然文の対訳文が参照される。そして、自然文(6)以外にも、翻訳対象語句“operation”及びコリレーションが高い単語“cancer”が含まれ、翻訳対象語句“operation”が対訳「手術」と対応付けられている自然文が対訳DBに登録されていることで、翻訳対象語句“operation”及びコリレーションが高い単語“cancer”が含まれている自然文における翻訳対象語句“operation”の高頻度の対訳は「手術」とであると判断され、翻訳対象語句“operation”及びコリレーションが高い単語“cancer”が含まれており、翻訳対象語句が判断した高頻度の対訳と対応付けられている自然文(6)が選択され、自然文(6)と対応付けられて対訳DBに登録されている対訳文(6)が、以下のように翻訳対象語句及びコリレーションの高い単語の対訳が強調表示されて出力される。

「ジョンさんの手術後、彼の癌は治った。」

従って、この例でも翻訳対象語句“operation”を原文と同じ意味で用いている適切な自然文(対訳文)が選択される。

- [0078] なお、図3では説明を簡単にするために“operation”を含む自然文の数を抑制しているが、実際にはより多くの自然文が登録されており、例えば処理対象の原文が翻訳対象語句“operation”を「手術」の意味で用いている文である場合には、コリレーションの高い単語として、処理対象の原文の内容に応じて“injury”(けが)、“hospital”(病院)等の単語が抽出され、例えば処理対象の原文が翻訳対象語句“operation”を「稼働」の意味で用いている文である場合には、コリレーションの高い単語として、処理対象の原文の内容に応じて“system”(システム)、“computer”(コンピュータ)等の単語が抽出され、例えば処理対象の原文が翻訳対象語句“operation”を「適用」の意味で用いている文である場合には、コリレーションの高い単語として、処理対象の原文の内

容に応じて“low”(法)等の単語が抽出されることになる。

- [0079] 一方、翻訳対象語句とのコリレーションの高い単語が抽出されなかった場合には、ステップ116の判定が否定されてステップ120へ移行し、ステップ120～134において、処理対象の原文中の翻訳対象語句以外の着目単語と代替可能な代替単語(請求項12に記載の代替語句に相当)に基づいて自然文を選択するスキーム解析を行う。
- [0080] すなわち、ステップ120では処理対象の原文中に存在し、かつステップ102の検索により対訳DBから抽出された各自然文中に存在しない単語(着目単語と称する)を判断する。ステップ122では、対訳DBに登録されている自然文の中から着目単語が含まれている自然文を検索する。なお、着目単語が複数になる場合もあるが、この場合は各着目単語についてステップ122の検索を行う。またステップ124では、ステップ122の検索により抽出された各自然文中の着目単語のみが相違している自然文(請求項13に記載の「着目語句を含む自然文の検索によって抽出された自然文と同一の構文の自然文」に相当)を各々検索する検索条件を設定し、設定した検索条件に基づき対訳DBに登録されている自然文の中から該当する自然文の検索を行う。
- [0081] 次のステップ126では、ステップ124の検索によって該当する自然文が抽出されたか否か判定する。着目単語が別の単語に入れ替わっている以外は同一の自然文が対訳DBに各々存在存在していた場合、前記別の単語は着目単語に代えて使用可能な代替単語である可能性が高い。このため、判定が肯定された場合はステップ128へ移行し、検索により抽出された自然文において、着目単語と置き換わっている単語を着目単語の代替単語と認識する。なお、上述したステップ120～130は請求項12(詳しくは請求項13)に記載の判断手段に対応している。
- [0082] そしてステップ130では、一致度最大(これに代えて「一致度所定値以上」としてもよい)の複数の自然文の中に着目単語が代替単語に置き換わっている自然文が有るか否か判定する。一致度最大の複数の自然文の中に、処理対象の原文中に存在する着目単語に代えて代替単語が含まれている自然文が存在していた場合、当該自然文は翻訳対象語句を処理対象の原文と同じ意味で用いている文である可能性が高いと判断できる。このため、判定が肯定された場合はステップ132へ移行し、一致

度最大(これに代えて「一致度所定値以上」としてもよい)でかつ着目単語に代えて代替単語が含まれている自然文と対応付けられて対訳DBに登録されている目的言語の対訳文を読み出し、読み出した対訳文上での翻訳対象語句及び代替単語の対訳を認識し、認識された翻訳対象語句の対訳が強調表示されると共に、代替単語の対訳も代替単語であることが識別可能にマーキングされて表示されるように、読み出した対訳文をディスプレイ12に表示して処理を終了する。この場合もユーザは指定した翻訳対象語句の適切な対訳(原文から目的言語の文として自然な対訳文が得られる可能性の高い対訳)を認識することができる。なお、上述したステップ130、132は請求項12に記載の第1の対訳判断手段に対応している。

[0083] なお、ステップ132においても、ステップ102の検索によって抽出された自然文のうち、一致度最大でかつ着目単語に代えて代替単語が含まれている自然文以外に、一致度の降順に複数の自然文(一致度の降順に一定数の自然文、或いは一致度が所定値以上の全ての自然文)の対訳文も対訳DBから読み出し、一致度及び不一致単語数に応じた順序でディスプレイ12に併せて一覧表示する。

[0084] (例6)上記処理について実例を挙げて更に説明する。原言語が英語、目的言語が日本語であり、“have lunch”を含む処理対象の原文における“have”が翻訳対象語句として指定されたが、対訳DBには“have breakfast”を含む自然文(翻訳対象語句“have”を処理対象の原文と同じ意味で用いている自然文)は登録されているものの、“have lunch”を含む自然文は登録されておらず、一致度に基づく判定やコリレーション解析によっても適正な対訳文を特定できなかった場合には、処理対象の原文における翻訳対象語句(“have”)や翻訳対象語句とのコリレーションが高いと判断した単語以外の各単語(例えば“lunch”等)が着目単語とされ、各着目単語が含まれている自然文が検索される。これにより、例えば着目単語“lunch”については、例えば“eat lunch”や“take a late lunch”を含む自然文が抽出される。

[0085] 次に、前記検索によって抽出された各自然文について、個々の自然文中の着目単語のみが相違している自然文が各々検索される。これにより、例えば着目単語“lunch”を含む自然文として抽出された自然文については、着目単語のみが相違している自然文として、例えば“eat breakfast”や“take a late breakfast”を含む自然文

が抽出される。従って、単語“breakfast”は着目単語“lunch”と代替可能である可能性の高い代替単語であると判断できるので、最初の検索(翻訳対象語句“have”を含む自然文の検索)で抽出された自然文のうち、着目単語“lunch”が代替単語“breakfast”に入れ替わっている自然文(“have breakfast”を含む自然文)と対応付けられて対訳DBに登録されている対訳文が、翻訳対象語句“have”の対訳(「食べる」)が強調表示されると共に、代替単語“breakfast”の対訳(「朝食」)がマーキングされて出力される。従って、この例でも翻訳対象語句“have”を処理対象の原文と同じ意味で用いている適切な自然文(対訳文)が選択される。

- [0086] また、同様に“have cats”を含む処理対象の原文における“have”が翻訳対象語句として指定されたが、対訳DBには“have dogs”を含む自然文は登録されているものの、“have cats”を含む自然文は登録されておらず、一致度に基づく判定やコリレーション解析によっても適正な対訳文を特定できなかった場合にも、“cats”等の単語が着目単語とされ、上記と同様にして単語“dogs”が着目単語“cats”と代替可能である可能性の高い代替単語であると判断されることで、最初の検索(翻訳対象語句“have”を含む自然文の検索)で抽出された自然文のうち、着目単語“cats”が代替単語“dogs”に入れ替わっている自然文(“have dogs”を含む自然文)と対応付けられて対訳DBに登録されている対訳文が、翻訳対象語句“have”の対訳(「飼う」)が強調表示されると共に、代替単語“dogs”の対訳(「犬」)がマーキングされて出力されることになる。従って、この例でも翻訳対象語句“have”を処理対象の原文と同じ意味で用いている適切な自然文(対訳文)が選択される。

- [0087] (例7)原言語が日本語、目的言語が英語であり、処理対象の原文「お湯をかける」における「かける」が翻訳対象語句として指定されたが、対訳DBには「水をかける」という自然文(翻訳対象語句「かける」を処理対象の原文と同じ意味で用いている自然文)は登録されているものの、「お湯をかける」は登録されておらず、例えば「電話をかける」等、他の自然文も登録されているために、一致度に基づく判定やコリレーション解析によっても適正な対訳文を特定できなかった場合には、処理対象の原文における「お湯」が着目単語とされ、着目単語が含まれている自然文が検索される。これにより、例えば「お湯につける」「お湯をわかす」「お湯を流す」「お湯で洗う」「お湯から引

き上げる」「お湯で温める」等の自然文が抽出される。

- [0088] 次に、抽出された上記各自然文について、個々の自然文中の着目単語のみが相違している自然文が各々検索される。そして、この検索の結果、上記各自然文について、着目単語「お湯」と代替可能な代替単語として、各々以下のような単語が得られる。

「一につける。」 …「ソース」「しょうゆ」「水」「スーツケース」「腕」

「一わかす。」 …無し

「一を流す。」 …「電気」「水」「番組」「うわさ」

「一で洗う。」 …「シャンプー」「スポンジ」「塩水」「水」「石鹸」

「一から引き上げる。」…「水」「海底」「会議」「ゼロ」

「一で温める。」 …「電子レンジ」「余熱」

上述した各単語のうち単語「水」は出現回数が多く、また「水をかける」という自然文も存在しているので、処理対象の原文「お湯をかける」における着目単語「お湯」と代替可能である可能性の高い代替単語であると判断できる。このため、最初の検索(翻訳対象語句“かける”を含む自然文の検索)で抽出された自然文のうち、着目単語「お湯」が代替単語「水」に入れ替わっている自然文「水をかける」と対応付けられて対訳DBに登録されている対訳文が、以下のように、翻訳対象語句「かける」の対訳(“pour”)が強調表示されると共に、代替単語「水」の対訳(“water”)がマーキングされて出力される。

“pour *water* into 一”

上記の対訳文において、“*…*”は単語“…”が代替単語であることを表している。従って、この例でも翻訳対象語句「かける」を処理対象の原文と同じ意味で用いている適切な自然文(対訳文)が選択される。

- [0089] また、同様に処理対象の原文「3をかける」における「かける」が翻訳対象語句として指定されたが、対訳DBには「4をかける」という自然文(翻訳対象語句「かける」を処理対象の原文と同じ意味で用いている自然文)は登録されているものの、「3をかける」は登録されておらず、一致度に基づく判定やコリレーション解析によっても適正な対訳文を特定できなかった場合には、「3」等の単語が着目単語とされ、上記と同様にし

て単語「4」が着目単語「3」と代替可能である可能性の高い代替単語であると判断されることで、最初の検索(翻訳対象語句「かける」を含む自然文の検索)で抽出された自然文のうち、着目単語「3」が代替単語「4」に入れ替わっている自然文「4をかける」と対応付けられて対訳DBに登録されている対訳文が、翻訳対象語句「かける」の対訳("multiply")が強調表示されると共に、代替単語「4」の対訳("four")がマーキングされて出力されることになる。

"multiply by *four*"

従って、この例でも翻訳対象語句「かける」を処理対象の原文と同じ意味で用いている適切な自然文(対訳文)が選択される。

[0090] 一方、着目単語のみが相違している自然文の検索において該当する自然文が抽出されなかった場合(ステップ126の判定が否定された場合)、又は、一致度最大の複数の自然文の中に着目単語が代替単語に置き換わっている自然文が存在していなかった場合(ステップ130の判定が否定された場合)には、最適な対訳文を特定することは困難であるため、ステップ134へ移行し、ステップ102の検索によって抽出された全ての自然文、或いは一致度の降順に複数の自然文(一致度の降順に一定数の自然文、或いは一致度が所定値以上の全ての自然文)と対応付けられて対訳DBに登録されている対訳文を各々読み出し、読み出した個々の対訳文上での翻訳対象語句の対訳を認識し、認識した翻訳対象語句の対訳を強調表示させた状態で、個々の対訳文を一致度及び不一致単語数に応じた順序で一致度最大の各自然文一致度の降順にディスプレイ12に一覧表示して処理を終了する。この場合、ユーザは指定した翻訳対象語句についての幾つかの適切な対訳候補を認識することができる。

[0091] なお、上記では一致単語数のみを用いて一致度を演算する態様を例に説明したが、これに限定されるものではなく、自然文と原文との一致単語数が多くなるに従って一致度が高くなり、かつ自然文と原文との不一致単語数が少なくなるに従って一致度が高くなるように一致度の演算式を定め、一致単語数及び不一致単語数に応じて一致度を演算・評価するようにしてもよい。この態様は請求項5記載の発明に対応している。また、当初は一致単語数のみに基づいて一致度の評価を行い、コリレーショ

ン解析やスキーム解析を行っても単一の自然文(対訳文)を選択することが困難な場合に不一致単語数を計数し、原文との不一致単語数が最小の自然文(の対訳文)を選択するようにしてもよい。

[0092] また、一致度の評価に際し、上記の一致単語数や不一致単語数以外に、自然文と原文との単語の並び順の類似度、或いは原文との一致単語の間に存在する自然文中の不一致単語数も併用し、自然文と原文との単語の並び順の類似度が高くなるに従って原文との一致度が高くなるように一致度を評価したり、一致単語の間に存在する不一致単語数が少なくなるに従って原文との一致度が高くなるように一致度を評価するようにしてもよい。自然文と原文との単語の並び順の類似度を用いる態様は請求項9記載の発明に対応しており、一致単語の間に存在する不一致単語数を用いることは請求項10記載の発明に対応している。自然文と原文との単語の並び順の類似度、及び原文との一致単語の間に存在する自然文中の不一致単語数も考慮した一致度の評価は、具体的には、例えば以下の処理によって実現することができる。

[0093] すなわち、まず原文中の各単語に、原文上での翻訳対象語句との隔たり(単語数)に応じて第1評価値を付与する。この第1評価値は、原文中の翻訳対象語句に対する第1評価値が最大となり、原文中の翻訳対象語句以外の各単語の第1評価値については、原文上での翻訳対象語句との隔たりが大きくなるに従って(間に存在している単語数が多くなるに従って)値が小さくなるように定めることができる。例えば原文が単語Aー単語Jの9個の単語から成り(原文=(A,B,C,D,E,F,G,H,J))、原文中の4番目の単語Dが翻訳対象語句として指定された場合には、原文中の各単語Aー単語Jに対して各々以下のような第1評価値を付与することができる。

A=1.0, B=2.0, C=5.0, D=10.0, E=5.0, F=2.0, G=1.0, H=0.5, J=0.2

[0094] 次に、ステップ102の検索により対訳DBから抽出された翻訳対象語句を含む各自然文について、自然文の各単語のうち原文中の何れかの単語と一致する一致単語を認識し、認識した各一致単語に対して先に付与した第1評価値の合計値を演算する。そして、この第1評価値の合計値を一致度とし、検索により抽出された各自然文の対訳文を、対応する各自然文の一致度(第1評価値の合計値)の降順に出力する。例えば検索により以下の単語配列の自然文1ー自然文5が抽出された場合(但し単

語Xは任意の単語)、

自然文1=(D,X,X,E,F,G)

自然文2=(D,X,E,F,G)

自然文3=(D,E,F,G)

自然文4=(A,B,X,C,X,D)

自然文5=(D,E,F,G,H,J)

各自然文の第1評価値の合計値は、

自然文1:10.0(単語D)+0(単語X)+0(単語X)+5.0(単語E)+2.0(単語F)+1.0(単語G)=18.0

自然文2:10.0(単語D)+0(単語X)+5.0(単語E)+2.0(単語F)+1.0(単語G)=18.0

自然文3:10.0(単語D)+5.0(単語E)+2.0(単語F)+1.0(単語G)=18.0

自然文4:1.0(単語A)+2.0(単語B)+0(単語X)+5.0(単語C)+0(単語X)+10.0(単語D)=18.0

自然文5:10.0(単語D)+5.0(単語E)+2.0(単語F)+1.0(単語G)+0.5(単語H)+0.2(単語J)=18.7

となり、第1評価値の合計値(一致度)は自然文5が最大となる。

[0095] 第1評価値は、原文中の翻訳対象語句以外の各単語に対し、原文上での翻訳対象語句との隔たりが小さくなるに従って値が大きくなるように定めているので、上記のように、翻訳対象語句を含む各自然文を第1評価値の合計値に基づいて評価することで、原文上で翻訳対象語句に近い位置に存在している単語をより多く含む自然文、すなわち原文中の翻訳対象語句及びその近辺の単語から成るフレーズ(連語)を含んでいる可能性の高い自然文(原文との単語の並び順の類似度が高いと推定される自然文)を、より一致度の高い自然文として評価することができる。

[0096] また、上記の例では自然文1～自然文4の一致度(第1評価値の合計値)が同値となっているが、このように第1評価値に基づく一致度が同値の自然文が複数存在していた場合には、自然文の各単語のうち原文中の何れかの単語と一致する一致単語に対し、各自然文上での翻訳対象語句との隔たり(単語数)に応じて第2評価値を付与することを、第1評価値に基づく一致度が同値の各自然文について各々行う。この

第2評価値についても、各自然文に含まれる翻訳対象語句に対する第2評価値が最大となり、各自然文のうち翻訳対象語句以外の一致単語の第2評価値については、各自然文上での翻訳対象語句との隔たりが大きくなるに従って(間に存在している単語数が多くなるに従って)値が小さくなるように定めることができる。そして、第1評価値に基づく一致度が同値の各自然文について各一致単語に付与した第2評価値の合計値を各々演算し、第1評価値に基づく一致度が同値の各自然文の対訳文を、対応する各自然文の第2評価値の合計値(これも本発明に係る一致度に含まれる)の降順に出力する。

- [0097] 例えば前出の自然文1ー自然文4については、各一致単語に対して以下のような第2評価値が付与され、以下のような第2評価値の合計値が得られる。なお、以下では、翻訳対象語句の第2評価値を10.0に設定すると共に、他の一致単語については、翻訳対象語句との間に存在する単語数が0,1,2,3,4…と増加していくに従って第2評価値が5.0,2.0,1.0,0.5,0.2…と減少するように第2評価値を設定した例を示している。

自然文1: $10.0(\text{単語D}) + 0(\text{単語X}) + 0(\text{単語X}) + 1.0(\text{単語E}) + 0.5(\text{単語F}) + 0.2(\text{単語G}) = 11.7$

自然文2: $10.0(\text{単語D}) + 0(\text{単語X}) + 2.0(\text{単語E}) + 1.0(\text{単語F}) + 0.5(\text{単語G}) = 13.5$

自然文3: $10.0(\text{単語D}) + 5.0(\text{単語E}) + 2.0(\text{単語F}) + 1.0(\text{単語G}) = 18.0$

自然文4: $0.2(\text{単語A}) + 0.5(\text{単語B}) + 0(\text{単語X}) + 2.0(\text{単語C}) + 0(\text{単語X}) + 10.0(\text{単語D}) = 12.7$

従って、自然文1ー自然文4(の対訳文)は、第2評価値の合計値(一致度)の降順、すなわち自然文3,2,4,1の順に出力される。

- [0098] 第2評価値は、各自然文の各単語のうち原文中の何れかの単語と一致する一致単語に対し、各自然文上での翻訳対象語句との隔たりが小さくなるに従って値が大きくなるように定めているので、上記のように、翻訳対象語句を含む各自然文を第2評価値の合計値に基づいて評価することで、原文との一致単語をより多く含みかつ一致単語の間に存在する不一致単語がなるべく少ない自然文、すなわち原文中の翻訳対象語句及びその近辺の単語から成るフレーズ(連語)を含んでいる可能性の高い

自然文を、より一致度の高い自然文として評価することができる。

[0099] また、自然文と原文との単語の並び順の類似度、及び原文との一致単語の間に存在する自然文中の不一致単語数も考慮した一致度の評価は、例えば以下の処理によって実現することも可能である。

[0100] すなわち、まず原文上で翻訳対象語句よりも前側に存在している前側単語群から翻訳対象語句との隔たりが最小の前側単語(このときは翻訳対象語句と隣り合っている前側単語)を抽出すると共に、原文上で翻訳対象語句よりも後側に存在している後側単語群から翻訳対象語句との隔たりが最小の後側単語(このときは翻訳対象語句と隣り合っている後側単語)を抽出する。そして、検索により対訳DBから抽出された翻訳対象語句を含む各自然文に対し、各自然文において翻訳対象語句が存在している位置を基準位置とし(自然文中の複数箇所に翻訳対象語句が各々存在している場合は何れかの翻訳対象語句が存在している位置を基準位置とする)、先に抽出した前側単語が基準位置から前側へ所定単語数の範囲内(例えば3単語以内)に存在しているか否かを探索すると共に、先に抽出した後側単語が基準位置から後側へ所定単語数の範囲内(例えば3単語以内)に存在しているか否かを探索する。そして、上記の探索によって前側単語及び後側単語が発見された自然文に対し、基準位置と前側単語及び後側単語の隔たりが小さくなるに従って値が大きくなるように定めた第3評価値を付与する。

[0101] 続いて、原文中の前側単語群から、未抽出でかつ翻訳対象語句との隔たりが最小の前側単語(このときは翻訳対象語句との間に単一の単語(前回の探索に用いた前側単語)が存在している前側単語)を抽出すると共に、原文中の後側単語群から、未抽出でかつ翻訳対象語句との隔たりが最小の後側単語(このときは翻訳対象語句との間に単一の単語(前回の探索に用いた後側単語)が存在している後側単語)を抽出する。そして、前回の探索によって前側単語及び後側単語が発見された各自然文に対し、先に抽出した前側単語が前回の探索で発見された前側単語の位置から前側へ所定単語数の範囲内(例えば3単語以内)に存在しているか否かを探索すると共に、先に抽出した後側単語が前回の探索で発見された後側単語の位置から後側へ所定単語数の範囲内(例えば3単語以内)に存在しているか否かを探索する。そし

て、今回の探索によって前側単語及び後側単語の少なくとも一方が発見された自然文に対し、前回の探索で発見された前側単語又は後側単語の位置と今回発見された前側単語又は後側単語の位置との隔たりが小さくなるに従って値が大きくなるように定めた第3評価値を付与する。

- [0102] なお、上記の第3評価値としては、上記のように、前回の探索で発見された前側単語又は後側単語の位置と今回発見された前側単語又は後側単語の位置との自然文上での隔たりのみに応じて一意に定まる評価値を用いてもよいが、原文上での翻訳対象語句と探索対象の前側単語又は後側単語との隔たりも考慮し、探索対象の前側単語又は後側単語が、原文上での翻訳対象語句との隔たりの小さな単語である程第3評価値に及ぼす影響が大きくなるように（探索対象の前側単語又は後側単語が、原文上での翻訳対象語句との隔たりが大きい単語になるに従って、付与される第3評価値が全体的に小さくなるように）、第3評価値を定めてもよい。また、上記のように、前回の探索で発見された前側単語又は後側単語の位置と今回発見された前側単語又は後側単語の位置との自然文上での隔たりが小さくなるに従って値が大きくなるように第3評価値を定めることに代えて、前述の基準位置と今回発見された前側単語又は後側単語の位置との自然文上での隔たりが小さくなるに従って第3評価値が大きくなるように第3評価値を定めてもよい。

- [0103] 以上の処理を、原文から翻訳対象語句との隔たりの昇順に前側単語及び後側単語を抽出しながら、原文から抽出可能な前側単語及び後側単語が無くなる迄繰り返した後に、各自然文に付与した第3評価値の合計値を演算し、この第3評価値の合計値を一致度とし、検索により抽出された各自然文の対訳文を、対応する各自然文の一致度（第3評価値の合計値）の降順に出力する。

- [0104] 例えば原文が単語A～単語Eの5個の単語から成り（原文＝（A,B,C,D,E））、原文中の3番目の単語Cが翻訳対象語句として指定された場合に、まず前側単語として単語Bが、後側単語として単語Dが原文から各々抽出され、検索により対訳DBから抽出された翻訳対象語句を含む各自然文に対し、前側単語Bが翻訳対象語句から前側へ所定単語数の範囲内（例えば3単語以内）に存在しているか否か、及び、後側単語Dが翻訳対象語句から後側へ所定単語数の範囲内（例えば3単語以内）に存

在しているか否かが各々探索され、前側単語B及び後側単語Dが発見された自然文に対して第3評価値が付与される。続いて、前側単語として単語Aが、後側単語として単語Eが原文から各々抽出され、前回の探索で前側単語B及び後側単語Dが発見された自然文に対し、前側単語Aが前側単語Bから前側へ所定単語数の範囲内(例えば3単語以内)に存在しているか否か、及び、後側単語Eが後側単語Dから後側へ所定単語数の範囲内(例えば3単語以内)に存在しているか否かが各々探索され、前側単語A及び後側単語Eが発見された自然文に対して第3評価値が付与される。

[0105] 以下に示す10個の自然文a〜自然文kに対して上記の処理を行って第3評価値の合計値の降順に並べたとすると、以下の順序となる。なお、以下の例において、単語Xは任意の単語、「/」は区切記号を意味している。

1. 自然文a=(A,B,C,D,E)
2. 自然文b=(A,B,C,D,/ ,B,C,D,E)
3. 自然文c=(B,C,D)
4. 自然文d=(A,B,C,X,D,E)
5. 自然文e=(A,B,C,X,D)
6. 自然文f=(A,B,C,/ ,C,D,E)
7. 自然文g=(A,X,B,C,X,D)
8. 自然文h=(B,C,/ ,C,D)
9. 自然文j=(B,X,C,/ ,C,X,D)
10. 自然文k=(B,X,X,C,/ ,C,X,X,D)

翻訳対象語句を含む各自然文に対し、上記のように第3評価値を付与し、第3評価値の合計値に基づいて評価することで、原文との単語の並び順の類似度が高く、かつ原文との一致単語の間に存在する不一致単語の数なるべく少ない自然文、すなわち原文が不一致単語をなるべく含まない状態(塊度の高い状態)で含まれている自然文を、より一致度の高い自然文として評価することができる。

[0106] 更に、自然文と原文との単語の並び順の類似度、及び原文との一致単語の間に存在する自然文中の不一致単語数も考慮した一致度の評価は、例えば以下の処理に

よって実現することもできる。

[0107] すなわち、まず原文上で翻訳対象語句よりも前側に存在している前側単語群から翻訳対象語句との隔たりが最小の前側単語(このときは翻訳対象語句と隣り合っている前側単語)を抽出し、検索により対訳DBから抽出された翻訳対象語句を含む各自然文に対し、各自然文において翻訳対象語句が存在している位置を基準位置とし(自然文中の複数箇所に翻訳対象語句が各々存在している場合は何れかの翻訳対象語句が存在している位置を基準位置とする)、先に抽出した前側単語が基準位置から前側へ所定単語数の範囲内(例えば3単語以内)に存在しているか否かを探索する。この処理は、翻訳対象語句を含む各自然文について、第4評価値として距離別単語数(詳しくは第1の距離別単語数及び第2の距離別単語数)を各々計数するものであり、上記の探索によって前側単語が発見された自然文に対し、基準位置と前側単語との距離(詳しくは基準位置と前側単語との間に存在する不一致単語数)を計数し、計数した距離に対応する第1の距離別単語数を各々カウントアップする。

[0108] 続いて、原文中の前側単語群から、未抽出でかつ翻訳対象語句との隔たりが最小の前側単語(このときは翻訳対象語句との間に単一の単語(前回の探索に用いた前側単語)が存在している前側単語)を抽出し、前回の探索によって前側単語が発見された各自然文に対し、先に抽出した前側単語が前回の探索で発見された前側単語の位置から前側へ所定単語数の範囲内(例えば3単語以内)に存在しているか否かを探索し、今回の探索によって前側単語が発見された自然文に対し、前回の探索で発見された前側単語の位置と今回発見された前側単語との距離を計数し、計数した距離に前回の探索で発見された前側単語と基準位置との距離を加算することで、今回の探索により発見した前側単語と基準位置との距離を求め、求めた距離に対応する第1の距離別単語数を各々カウントアップする。この処理を、原文から翻訳対象語句との隔たりの昇順に前側単語を抽出しながら、原文から抽出可能な前側単語が無くなる迄繰り返す。

[0109] 例えば原文が単語Aー単語Eの5個の単語から成り(原文=(A,B,C,D,E))、原文中の4番目の単語Dが翻訳対象語句として指定された場合、まず前側単語として単語Cが原文から抽出され、検索により対訳DBから抽出された翻訳対象語句を含む各

自然文に対し、前側単語Bが翻訳対象語句から前側へ所定単語数の範囲内(例えば3単語以内)に存在しているか否かが各々探索され、前側単語Cが発見された自然文に対して第1の距離別単語数のカウントアップが行われる。この時点での第1の距離別単語数の計数結果の一例を次の表1に示す。

[0110] [表1]

〈 単語Cを探索した段階での第1の距離別単語数の計数結果の一例 〉

自然文の一例	距離別単語数			
	距離=0	1	2	3
CD	1	0	0	0
CXD	0	1	0	0
CXXD	0	0	1	0

[0111] 続いて、前側単語として単語Bが原文から抽出され、前回の探索で前側単語Cが発見された自然文に対し、前側単語Cから前側へ所定単語数の範囲内(例えば3単語以内)に前側単語Bが存在しているか否かが各々探索され、前側単語Bが発見された自然文に対して第1の距離別単語数のカウントアップが行われる。この時点での第1の距離別単語数の計数結果の一例を次の表2に示す。

[0112] [表2]

〈 単語Bを探索した段階での第1の距離別単語数の計数結果の一例 〉

自然文の一例	距離別単語数				
	距離=0	1	2	3	4
BCD	2	0	0	0	0
BXCD	1	1	0	0	0
BXXCD	1	0	1	0	0
BCXD	0	2	0	0	0
BXCXD	0	1	1	0	0
BXXCXD	0	1	0	1	0
BCXXD	0	0	2	0	0
BXCXXD	0	0	1	1	0
BXXCXXD	0	0	1	0	1

[0113] 更に、前側単語として単語Aが原文から抽出され、前回の探索で前側単語Bが発

見された自然文に対し、前側単語Bから前側へ所定単語数の範囲内(例えば3単語以内)に前側単語Aが存在しているか否かが各々探索され、前側単語Aが発見された自然文に対して第1の距離別単語数のカウントアップが行われる。この時点での第1の距離別単語数の計数結果の一例を次の表3に示す。

[0114] [表3]

＜ 単語Aを探索した段階での第1の距離別単語数の計数結果の一例 ＞

自然文の一例	距離別単語数						
	距離=0	1	2	3	4	5	6
ABCD	3	0	0	0	0	0	0
AXBCD	2	1	0	0	0	0	0
AXXBCD	2	0	1	0	0	0	0
ABXCD	1	2	0	0	0	0	0
AXBXCD	1	1	1	0	0	0	0
AXXBXCD	1	1	0	1	0	0	0
ABXXCD	1	0	2	0	0	0	0
AXBXXCD	1	0	1	1	0	0	0
AXXBXXCD	1	0	1	0	1	0	0
ABCXD	0	3	0	0	0	0	0
AXBCXD	0	2	1	0	0	0	0
AXXBCXD	0	2	0	1	0	0	0
ABXCXD	0	1	2	0	0	0	0
AXBXCXD	0	1	1	1	0	0	0
AXXBXCXD	0	1	1	0	1	0	0
ABXXCXD	0	1	0	2	0	0	0
AXBXXCXD	0	1	0	1	1	0	0
AXXBXXCXD	0	1	0	1	0	1	0
ABCXXD	0	0	3	0	0	0	0
AXBCXXD	0	0	2	1	0	0	0
AXXBCXXD	0	0	2	0	1	0	0
ABXCXXD	0	0	1	2	0	0	0
AXBXCXXD	0	0	1	1	1	0	0
AXXBXCXXD	0	0	1	1	0	1	0
ABXXCXXD	0	0	1	0	2	0	0
AXBXXCXXD	0	0	1	0	1	1	0
AXXBXXCXXD	0	0	1	0	1	0	1

[0115] 次に、検索により対訳DBから抽出された翻訳対象語句を含む各自然文に対し、上述した第1の距離別単語数とは別の方法で単語間の距離を計数し、第2の距離別単語

語数として集計する。具体的には、原文中の前側単語群のうち翻訳対象語句と隣り合っている前側単語が原文中に存在していないものと見なし(各自然文に含まれる単語Cを任意の不一致単語Xと見なし)、この状態で、原文中の前側単語群から翻訳対象語句との隔たりが最小の前側単語を抽出し、検索により対訳DBから抽出された翻訳対象語句を含む各自然文に対し、各自然文において翻訳対象語句が存在している位置を基準位置とし、抽出した前側単語が基準位置から前側へ所定単語数の範囲内(例えば3単語以内)に存在しているか否かを探索する。そして、上記の探索によって前側単語が発見された自然文に対し、基準位置と前側単語との距離(詳しくは基準位置と前側単語との間に存在する不一致単語数)を計数し、計数した距離に「1」を加えた距離を基準位置との距離とし、該基準位置との距離に対応する第2の距離別単語数を各々カウントアップする。

[0116] 続いて、原文中の前側単語群から、未抽出でかつ翻訳対象語句との隔たりが最小の前側単語を抽出し、前回の探索によって前側単語が発見された各自然文に対し、先に抽出した前側単語が前回の探索で発見された前側単語の位置から前側へ所定単語数の範囲内(例えば3単語以内)に存在しているか否かを探索し、今回の探索によって前側単語が発見された自然文に対し、前回の探索で発見された前側単語の位置と今回発見された前側単語との距離を計数し、計数した距離に前回の探索で発見された前側単語と基準位置との距離を加算することで、今回の探索により発見した前側単語と基準位置との距離を求め、求めた距離に対応する第2の距離別単語数を各々カウントアップする。この処理を、原文から翻訳対象語句との隔たりの昇順に前側単語を抽出しながら、原文から抽出可能な前側単語が無くなる迄繰り返す。

[0117] 例えば原文が単語Aー単語Eの5個の単語から成り(原文=(A,B,C,D,E))、原文中の4番目の単語Dが翻訳対象語句として指定された場合には、まず原文中の前側単語群のうち翻訳対象語句と隣り合っている前側単語Cが原文中に存在していないものと見なされ(各自然文に含まれる単語Cが任意の不一致単語Xと見なされ)た状態で、原文から単語B及び単語Aが順次抽出されて各自然文上での基準位置との距離が計数されることで、例として次の表4に示すような計数結果が得られる。

[0118] [表4]

＜ 第2の距離別単語数に係る距離計数結果の一例 ＞

自然文の一例	B D の距離 (実距離+1)	(A B の距離)	A D の距離 (BD+AB)
A B C D	2	0	2
A X B C D	2	1	3
A X X B C D	2	2	4
A B X C D	3	0	3

[0119] そして、上記の計数結果に基づいて第2の距離別単語数をカウントアップすることで、次の表5に示すような第2の距離別単語数の計数結果が得られる。

[0120] [表5]

＜ 第2の距離別単語数の計数結果の一例 ＞

自然文の一例	距離別単語数						
	距離=0	1	2	3	4	5	6
A B C D	0	0	2	0	0	0	0
A X B C D	0	0	1	1	0	0	0
A X X B C D	0	0	1	0	1	0	0
A B X C D	0	0	0	2	0	0	0

[0121] 上記のようにして、翻訳対象語句を含む各自然文に対し、原文中の前側単語群に関して第1の距離別単語数及び第2の距離別単語数の計数が完了すると、各自然文毎に、第1の距離別単語数及び第2の距離別単語数の計数結果を比較し、第1の距離別単語数と第2の距離別単語数のうち、単語間距離がより短いとの結果を示している方の計数結果を前側単語群に関する最終評価として選択する。具体的には、例えば自然文(A,B,C,D)は、表3に示す第1の距離別単語数では距離0で単語数3であるのに対し、表5に示す第2の距離別単語数では距離0で単語数0、距離1で単語数0、距離2で単語数2であるので、距離0における単語数の多い第1の距離別単語数が、自然文(A,B,C,D)の前側単語群に関する最終評価として選択される。上記の例は第1の距離別単語数及び第2の距離別単語数における距離0の単語数が相違し

ている場合であるが、距離0の単語数が同一の場合は距離1, 2, …の単語数が順次比較され、第1の距離別単語数及び第2の距離別単語数のうち同一距離の単語数がより多い方が最終評価として選択される。

[0122] なお、第2の距離別単語数の計数を行っているのは、原文に対して一部の単語の順序が入れ替わっている自然文についても原文との単語の並び順の類似度を正當に評価するためである。例えば自然文=(C,A,B,D)については、第1の距離別単語数では距離0で単語数0、距離1で単語数0、距離2で単語数1となるのに対し、第2の距離別単語数では距離0で単語数0、距離1で単語数2、距離2で単語数0となり、第2の距離別単語数が上記の自然文の前側単語群に関する最終評価として選択される。上記の例において、第1の距離別単語数の計数結果は自然文(C,A,B,D)を(C,X,X,D)と見なして評価しているのに対し、第2の距離別単語数の計数結果は自然文(C,A,B,D)を(X,A,B,(単語C抜け),D)と見なして評価しており、上記の自然文を、原文がより塊度の高い状態で含まれていると評価することができる。

[0123] 検索により対訳DBから抽出された翻訳対象語句を含む各自然文に対し、上記処理を行うことで前側単語群に関する最終評価が定まると、上記各自然文に対し、原文上で翻訳対象語句よりも後側に存在している後側単語群を用いて同様の処理(第1の距離別単語数及び第2の距離別単語数の計数、後側単語群に関する最終評価の決定)を行う。次に、各自然文毎に、得られた前側単語群に関する最終評価と後側単語群に関する最終評価を加算集計する(各最終評価における距離別単語数を同一距離毎に加算する)ことで、各自然文に対する総合評価を求める。そして、各自然文毎の総合評価に基づき、各自然文の対訳文を、対応する各自然文の総合評価が単語間距離がより短いとの結果を示している対訳文から順に出力する。

[0124] 具体的には、例えば各自然文を距離0における単語数の降順に並べ替えた後に、距離0における単語数が同値の自然文を距離1における単語数の降順に並べ替えることを、比較する単語数に対応する距離を1ずつインクリメントしながら繰り返すことで、総合評価が単語間距離がより短いとの結果を示している順に各自然文を並べ替える。そして、各自然文に対応する対訳文を順に出力する。

[0125] これにより、以下に示す9個の自然文a〜自然文jに対して上記の処理を行い、総合

評価に基づいて並べ替えを行ったとすると、以下の順序となる。

1. 自然文a=(A,B,C,D,E)
2. 自然文b=(A,B,C,D,/,B,C,D,E)
3. 自然文c=(A,B,C,X,D)
4. 自然文d=(B,C,D)
5. 自然文e=(A,X,B,C,X,D)
6. 自然文f=(B,C,X,D)
7. 自然文g=(B,C,/,C,D)
8. 自然文h=(B,X,C,/,C,X,D)
9. 自然文j=(B,X,X,C,/,C,X,X,D)

なお、上記処理において、第2の距離別単語数の計数は必須ではなく、例えば原文に対して一部の単語の順序が入れ替わっている自然文を評価する必要がない等の場合には、第2の距離別単語数の計数を省略し、第1の距離別単語数をそのまま最終評価として用いてもよい。

[0126] また、上述した第1評価値及び第2評価値／第3評価値／第4評価値(第1の距離別単語数及び第2の距離別単語数)に基づく一致度の評価・自然文の選択又は並べ替えは、コリレーション解析やスキーム解析による自然文の評価・選択又は並べ替えと併用できることは言うまでもないが、第1評価値と第2評価値又は第3評価値に基づいて自然文の一致度を評価する態様は、原文と自然文との一致単語数のみを用いて一致度を評価する態様等と比較して、原文中に存在する連語やフレーズ等、原文全体よりも狭い範囲の語句について参考となる対訳文を検索する等の場合により適しており、例えば通常の検索モード以外にフレーズ検索モード等の検索モードを設け、このフレーズ検索モードが選択された場合に第1評価値と第2評価値又は第3評価値に基づく一致度の評価・自然文の選択又は並べ替えを行うようにしてもよい。

[0127] また、上記ではまず一致度を演算して自然文の選択(対訳の判断)を行い、一致度では適切な自然文を絞り込めない場合にコリレーション解析による自然文の選択(対訳の判断)を行い、コリレーション解析でも適切な自然文を絞り込めない場合にスキーム解析による自然文の選択(対訳の判断)を行う例を説明したが、これに限定され

るものではなく、検索手段による検索によって抽出された翻訳対象語句を含む個々の自然文に対して一致度の演算、コリレーション解析及びスキーム解析を各々行うことで個々の自然文を評価し、評価結果に基づき、例えば次の表6に示すような優先順位に従って自然文の選択を行うようにしてもよい。

[0128] [表6]

〈 自然文選択の優先順位の一例 〉

一致度数	コリレーション解析 (高コリレーション単語の有無)	スキーム解析 (代替単語の有無)	
		有	無
高	有	1	
	無	2	5
低	有	3	
	無	4	6

[0129] 更に、個々の自然文に対して一致度の演算、コリレーション解析及びスキーム解析を各々行い、次式に従って総合評価点を演算し、総合評価点が最大の自然文を選択するようにしてもよい。

総合評価点 = $a \times (\text{一致度}) + b \times (\text{コリレーション評価点}) + c \times (\text{スキーム評価点})$

なお、上記式において、コリレーション評価点はコリレーション解析の結果に基づいて個々の自然文に付与した評価点、スキーム評価点はスキーム解析の結果に基づいて個々の自然文に付与した評価点、 a , b , c は予め定めた係数である。

[0130] また、先に説明した対訳判断処理(図2)では、ステップ102の検索によって処理対象の原文と完全一致している自然文が対訳DBから抽出されたか否かを判定し(ステップ104)、判定が肯定された場合は、処理対象の原文と完全一致している自然文の対訳文を読み出して表示する(ステップ106)ようにしていたが、これに限定されるものではなく、上記のステップ104, 106を省略し、ステップ102の検索によって処理対象の原文と完全一致している自然文が対訳DBから抽出されたか否かに拘らず、ステップ108以降の処理(一致度の演算等の処理)を無条件に行うようにしてもよい。この態様において、ステップ102の検索によって複数の自然文が抽出された場合にも、抽出された複数の自然文の中に原文と完全一致している自然文が含まれていれば、

該自然文の一致度が最大となるので、結果としてステップ104, 106を行った場合と同一の処理結果が得られる。

- [0131] また、本発明における対訳選択の精度は対訳DBに登録されている自然文及び対訳文の数に依存し、自然文及び対訳文の数が多くなるに従って対訳選択の精度は向上する。このため、本発明に係る対訳判断装置に、翻訳された原文と翻訳文を読み込み、読み込んだ原文と翻訳文をそのまま、或いは取捨選択した後に、自然文及び対訳文として対訳DBへ自動的に追加登録する機能を設けてもよい。本発明によれば、対訳DBには自然文及び対訳文を追加登録すればよく、自然文に対して単語単位での分割等の編集・加工を行う必要はないので、対訳DBへ情報(自然文及び対訳文)を自動的に追加登録することは極めて容易に実現できる。
- [0132] また、上記では対訳文の出力に際して翻訳対象語句の対訳を強調表示する態様を例に説明したが、これに限定されるものではなく、翻訳対象語句の対訳を強調表示しない態様も本発明の権利範囲に含まれることは言うまでもない。
- [0133] また、本発明における対訳選択の精度は対訳DBに登録されている自然文及び対訳文の内容の冗長度にも依存し、対訳DBに登録されている自然文及び対訳文の内容の冗長度が高くなると、対訳DBに登録されている自然文及び対訳文の数(対訳DBの容量)に比して対訳選択の精度は低下する。このため、本発明に係る対訳判断装置に、対訳DBに登録されている個々の自然文及び対訳文の中に内容の類似度が高い自然文及び対訳文が存在しているか否かを探索し、内容の類似度が非常に高い自然文対及び対訳文対を発見した場合には、一方の自然文及び対訳文を対訳DBから削除する機能を設けてもよい。
- [0134] 更に、上記では単一の対訳DBを用いて対訳の選択を行う例を説明したが、これに限定されるものではなく、例えば特定分野のマニュアルの翻訳に用いる等、処理対象の原文が特定の分野に限られる等の場合には、対訳DBを処理対象の原文の分野毎に分割するようにしてもよい。この場合、個々の対訳DBの容量の増大を抑制できると共に、自然文の検索に要する時間の短縮、対訳の選択精度の向上等の効果が得られる。
- [0135] また、上記ではユーザがPC10を介して行った翻訳対象対象語句の対訳の出力を

指示する所定の操作に従い、PC10で対訳判断処理が実行される態様を例を説明したが、これに限定されるものではなく、例として図4に示すように、インターネットやLAN(Local Area Network)等のネットワーク30にクライアントPC32及びサーバ・コンピュータ34が各々接続され、対訳DBを記憶するHDD等から成る記憶媒体36がサーバ・コンピュータ34に接続されたコンピュータ・システム38に本発明を適用し、ユーザからクライアントPC32を介してオンラインで対訳の問い合わせがあると、サーバ・コンピュータ34が対訳を判断してオンラインで回答するように構成することも可能である。

[0136] 図4のコンピュータ・システム38において、ユーザによる対訳の問い合わせ(図4の(1)参照)は、例えばユーザがクライアントPC32を介して原言語のテキストのデータをサーバ・コンピュータ34へ送信するか、又は原言語のテキストを含むウェブページのURL(Uniform Resource Locator)をサーバ・コンピュータ34へ通知することで、原言語のテキストを指定することによって行うことができる。この場合、サーバ・コンピュータ34は、記憶媒体36に記憶された対訳DBにアクセスしながら対訳判断処理を実行することで、指定されたテキストの対訳(文)を判断し(図4の(2)参照)、判断した対訳(文)をクライアントPC32へ送信することで問い合わせに回答する(図4の(3)参照)ように構成することができる。この構成は、サーバ・コンピュータ34の対訳判断機能(及び対訳DB)を複数のユーザが共有可能であるので、特に対訳DBの容量が大きい等の場合のコスト抑制に有効である。

[0137] また、上記では本発明を電子辞書装置に適用した態様を例に説明したが、これに限定されるものではなく、機械翻訳(自動翻訳)装置や自動通訳装置(音声で入力された原文を音声認識によって認識し、認識した原文の対訳文を判断し、判断した対訳文を音声で出力することで通訳を行う装置)、語学修得装置等にも適用可能であることは言うまでもない。図5には本発明を適用した機械翻訳装置40の概略構成の一例が示されている。図5において、原文入力部42は翻訳対象の原文(のテキストデータ)を翻訳対象語句選択部44及び対訳文組立部46に各々入力する。翻訳対象語句選択部44は、入力された翻訳対象の原文中の特定の語句を翻訳対象語句として選択し、選択した翻訳対象語句の対訳を対訳判断部48に問い合わせることを、翻訳対象語句として翻訳対象の原文中の個々の語句を順次選択しながら繰り返す。対訳

判断部48は本発明に係る対訳判断装置に相当する部分であり、記憶媒体50に記憶された対訳DBにアクセスしながら対訳判断処理を実行することで、問い合わせのあった翻訳対象語句の対訳を判断し、判断した対訳を対訳文組立部46へ出力する。

[0138] 対訳文組立部46は、翻訳対象語句選択部44から入力された翻訳対象語句と対訳判断部48から入力されるその対訳、及び原文入力部42から入力される翻訳対象の原文に基づき、対訳判断部48から入力される対訳を繋げる(必要に応じて順序を入れ替える)ことで、目的言語の対訳文を組み立てる。対訳文組立部46によって組み立てられた対訳文は対訳文出力部52へ出力され、ディスプレイ等への表示、記録媒体への記録、音声として出力する等の対訳文出力処理が行われる。上記構成では、翻訳対象語句の対訳の判断に本発明を適用しているので、翻訳対象語句の対訳として、原文から目的言語の文として自然な対訳文が得られる可能性の高い対訳が得られ、対訳文出力部52によって出力される対訳文が、目的言語の文として自然な対訳文になる。また、品詞判断や構文解析等の複雑な処理が不要となり、処理を簡略化することができる。なお、図5に示す機械翻訳装置においても、先に説明した図4のように、オンラインでの対訳文の問い合わせに対し、対訳の判断・組立等の処理を経て得られた対訳文をオンラインで回答するようにしてもよいことは言うまでもない。

[0139] 更に、上記では対訳DB(記憶手段)に対応付けて登録される一対の文(異なる言語で記述された一対の文)に対し、便宜的に一方を自然文、他方を対訳文と称していたが、第1の言語から第2の言語への翻訳(対訳の判断)のための作成された対訳DBを、第2の言語から第1の言語への翻訳(対訳の判断)に用いることも可能である。

[0140] また、上記では原言語や目的言語の一例として英語、日本語、中国語を挙げて説明したが、これに限定されるものではなく、原言語や目的言語として、例えばフランス語、ドイツ語、イタリア語、スペイン語、ポルトガル語等、任意の言語を適用可能であることは言うまでもない。

符号の説明

[0141] 10 PC

18 HDD

34 サーバ・コンピュータ

36 記憶媒体

40 機械翻訳装置

48 対訳判断部

50 記憶媒体

請求の範囲

- [1] 複数の単語から成る原言語の自然文を目的言語の対訳文と対応付けて複数記憶する記憶手段と、
前記記憶手段に記憶されている原言語の複数の自然文の中から、原言語の原文中の翻訳対象語句が含まれている自然文を検索する検索手段と、
前記検索手段による検索によって抽出された自然文と前記原文との一致度を求め、求めた一致度に基づいて選択した自然文の対訳文における少なくとも前記翻訳対象語句の対訳を、前記原文中の少なくとも前記翻訳対象語句の対訳と判断する第1の対訳判断手段と、
を含む対訳判断装置。
- [2] 前記検索手段は、前記記憶手段に記憶されている原言語の複数の自然文の中から、前記原文と完全一致している自然文も検索し、
前記第1の対訳判断手段は、前記検索手段による検索によって前記原文と完全一致している自然文が抽出された場合に、当該完全一致している自然文の対訳文を前記原文の対訳文と判断することを特徴とする請求項1記載の対訳判断装置。
- [3] 前記第1の対訳判断手段は、前記検索手段による検索によって抽出された自然文と前記原文との一致単語数を計数し、計数した一致単語数が多くなるに従って前記原文との一致度が高くなるように、前記原文との一致度を評価することを特徴とする請求項1記載の対訳判断装置。
- [4] 前記第1の対訳判断手段は、前記一致度として、前記計数した一致単語数を、前記翻訳対象語句を構成する単語数で除した値を求めることを特徴とする請求項3記載の対訳判断装置。
- [5] 前記第1の対訳判断手段は、前記検索手段による検索によって抽出された自然文と前記原文との不一致単語数も計数し、計数した不一致単語数が少なくなるに従って前記原文との一致度が高くなるように、前記原文との一致度を評価することを特徴とする請求項3記載の対訳判断装置。
- [6] 前記第1の対訳判断手段は、前記一致単語数の計数に際し、予め定められた頻出単語を計数対象から除外することを特徴とする請求項3記載の対訳判断装置。

- [7] 前記第1の対訳判断手段は、前記一致単語数又は前記不一致単語数の計数に際し、単複又は時制の相違に起因して語尾が相違している単語を一致単語とみなして計数することを特徴とする請求項3又は請求項5記載の対訳判断装置。
- [8] 前記第1の対訳判断手段は、前記一致単語数の計数に際し、複数回出現した一致単語を重複計数しないことを特徴とする請求項3記載の対訳判断装置。
- [9] 前記第1の対訳判断手段は、前記検索手段による検索によって抽出された自然文と前記原文との単語の並び順の類似度を評価し、単語の並び順の類似度が高くなるに従って前記原文との一致度が高くなるように、前記原文との一致度を評価することを特徴とする請求項3又は請求項5記載の対訳判断装置。
- [10] 前記第1の対訳判断手段は、前記検索手段による検索によって抽出された自然文において、前記原文との一致単語の間に存在する不一致単語数を計数し、計数した前記一致単語の間に存在する不一致単語数が少なくなるに従って前記原文との一致度が高くなるように、前記原文との一致度を評価することを特徴とする請求項3又は請求項5記載の対訳判断装置。
- [11] 前記検索手段による検索によって抽出された自然文に基づいて、前記翻訳対象語句と原言語の同一の文中に出現する頻度の高い頻出語句を認識する認識手段を更に備え、
前記第1の対訳判断手段は、前記検索手段による検索によって抽出された自然文のうち、前記認識手段によって認識されかつ前記原文中に存在する特定の頻出語句と、前記翻訳対象語句が各々含まれている自然文の対訳文を参照することで、前記翻訳対象語句及び前記特定の頻出語句が各々含まれる自然文の対訳文における翻訳対象語句についての高頻度の対訳を認識し、認識した高頻度の対訳を、前記原文中の前記翻訳対象語句の対訳と判断することを特徴とする請求項1記載の対訳判断装置。
- [12] 前記原文中に存在しかつ前記検索手段による検索によって抽出された自然文中に含まれていない着目語句と代替可能な代替語句を判断する判断手段を更に備え、
前記第1の対訳判断手段は、前記検索手段による検索によって抽出された自然文のうち、前記判断手段によって判断された代替語句及び前記翻訳対象語句が各々

含まれている自然文の対訳文における少なくとも前記翻訳対象語句の対訳を、前記原文中の少なくとも前記翻訳対象語句の対訳と判断することを特徴とする請求項1記載の対訳判断装置。

- [13] 前記判断手段は、前記着目語句を含む自然文を前記記憶手段に記憶されている複数の自然文の中から検索し、該検索によって抽出された自然文と同一の構文の自然文を前記記憶手段に記憶されている自然文の中から検索し、該検索によって抽出された自然文中で着目語句と置き換わっている語句を、前記代替語句と判断することを特徴とする請求項12記載の対訳判断装置。

- [14] 複数の単語から成る原言語の自然文を目的言語の対訳文と対応付けて複数記憶する記憶手段と、

前記記憶手段に記憶されている原言語の複数の自然文の中から、原言語の原文中の翻訳対象語句が含まれている自然文を検索する検索手段と、

前記検索手段による検索によって抽出された自然文に基づいて、前記翻訳対象語句と原言語の同一の文中に出現する頻度の高い頻出語句を認識する認識手段と、

前記検索手段による検索によって抽出された自然文のうち、前記認識手段によって認識されかつ前記原文中に存在する特定の頻出語句と、前記翻訳対象語句が各々含まれている自然文の対訳文を参照することで、前記翻訳対象語句及び前記特定の頻出語句が各々含まれる自然文の対訳文における翻訳対象語句についての高頻度の対訳を認識し、認識した高頻度の対訳を、前記原文中の前記翻訳対象語句の対訳と判断する第2の対訳判断手段と、

を含む対訳判断装置。

- [15] 前記翻訳対象語句が複数の単語から成る場合、前記検索手段は、前記記憶手段に記憶されている原言語の複数の自然文の中から、前記翻訳対象語句を構成する複数の単語のうちの少なくとも1つが含まれている自然文を検索することを特徴とする請求項1又は請求項14記載の対訳判断装置。

- [16] 目的言語の対訳文と対応付けて記憶手段に複数記憶されている、各々複数の単語から成る原言語の自然文の中から、原言語の原文中の翻訳対象語句が含まれている自然文を検索する第1のステップ、

及び、前記第1のステップにおける検索によって抽出された自然文と前記原文との一致度を求め、求めた一致度に基づいて選択した自然文の対訳文における少なくとも前記翻訳対象語句の対訳を、前記原文中の少なくとも前記翻訳対象語句の対訳と判断する第2のステップ

を含む対訳判断方法。

- [17] 目的言語の対訳文と対応付けて記憶手段に複数記憶されている、各々複数の単語から成る原言語の自然文の中から、原言語の原文中の翻訳対象語句が含まれている自然文を検索する第1のステップ、

前記第1のステップにおける検索によって抽出された自然文に基づいて、前記翻訳対象語句と原言語の同一の文中に出現する頻度の高い頻出語句を認識する第2のステップ、

前記第1のステップにおける検索によって抽出された自然文のうち、前記第2のステップで認識されかつ前記原文中に存在する特定の頻出語句と、前記翻訳対象語句が各々含まれている自然文の対訳文を参照することで、前記翻訳対象語句及び前記特定の頻出語句が各々含まれる自然文の対訳文における翻訳対象語句についての高頻度の対訳を認識し、認識した高頻度の対訳を、前記原文中の前記翻訳対象語句の対訳と判断する第3のステップ

を含む対訳判断方法。

- [18] 複数の単語から成る原言語の自然文を目的言語の対訳文と対応付けて複数記憶する記憶手段と接続されたコンピュータを、

前記記憶手段に記憶されている原言語の複数の自然文の中から、原言語の原文中の翻訳対象語句が含まれている自然文を検索する検索手段、

及び、前記検索手段による検索によって抽出された自然文と前記原文との一致度を求め、求めた一致度に基づいて選択した自然文の対訳文における少なくとも前記翻訳対象語句の対訳を、前記原文中の少なくとも前記翻訳対象語句の対訳と判断する第1の対訳判断手段

として機能させるプログラム。

- [19] 複数の単語から成る原言語の自然文を目的言語の対訳文と対応付けて複数記憶

する記憶手段と接続されたコンピュータを、

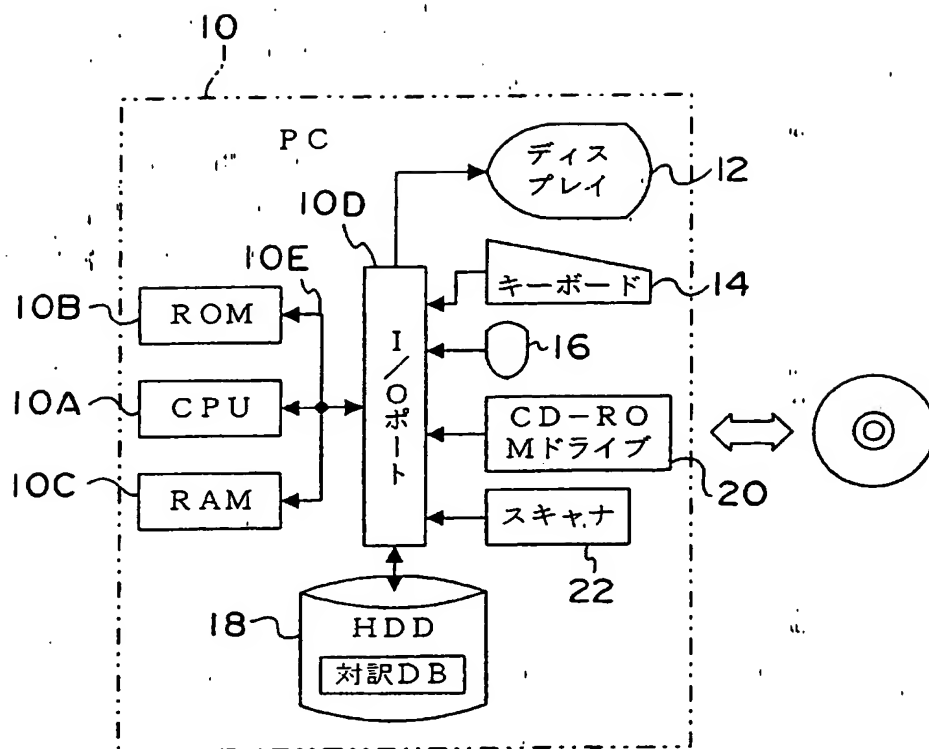
前記記憶手段に記憶されている原言語の複数の自然文の中から、原言語の原文中の翻訳対象語句が含まれている自然文を検索する検索手段、

前記検索手段による検索によって抽出された自然文に基づいて、前記翻訳対象語句と原言語の同一の文中に出現する頻度の高い頻出語句を認識する認識手段、

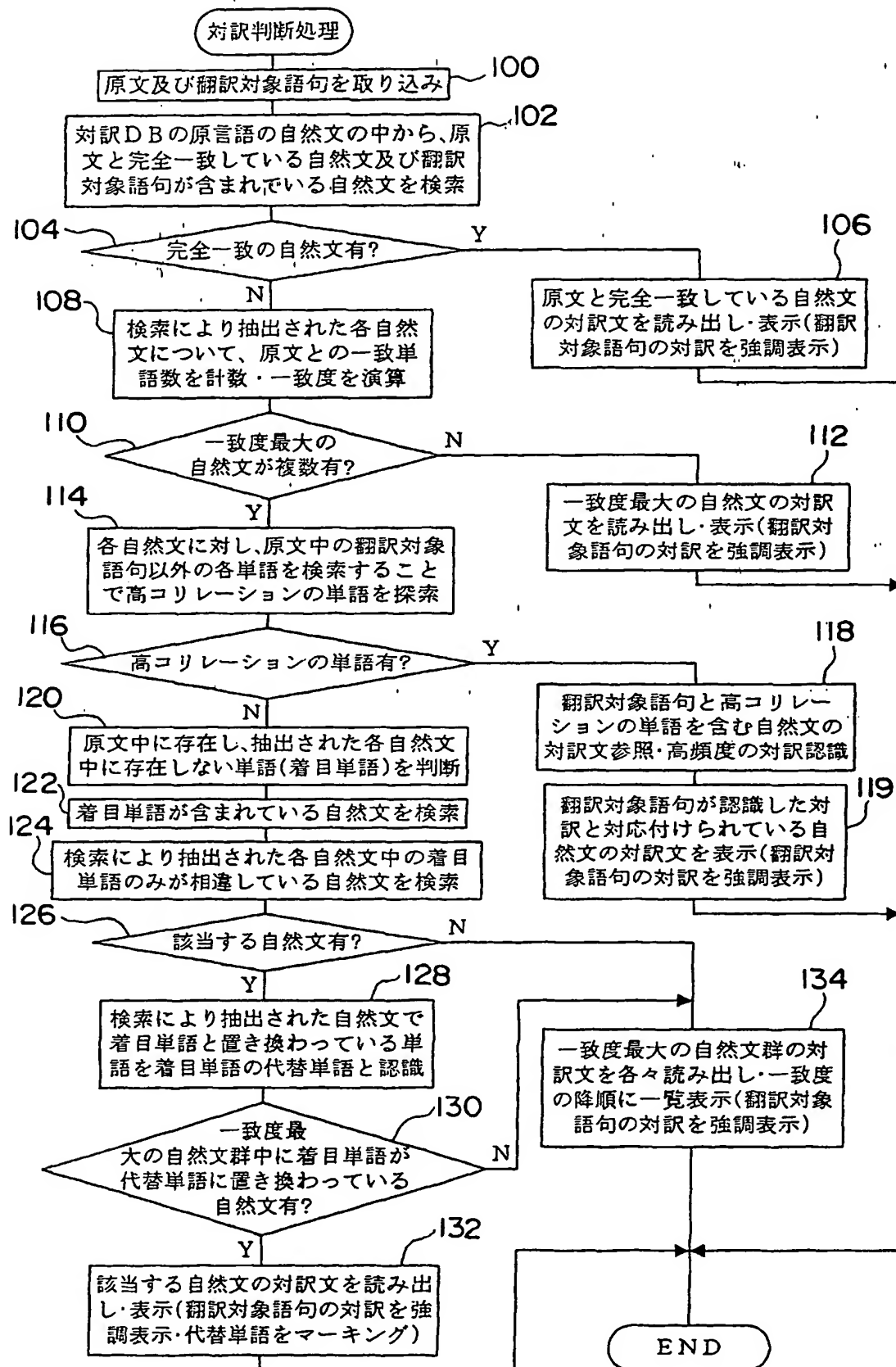
及び、前記検索手段による検索によって抽出された自然文のうち、前記認識手段によって認識されかつ前記原文中に存在する特定の頻出語句と、前記翻訳対象語句が各々含まれている自然文の対訳文を参照することで、前記翻訳対象語句及び前記特定の頻出語句が各々含まれる自然文の対訳文における翻訳対象語句についての高頻度の対訳を認識し、認識した高頻度の対訳を、前記原文中の前記翻訳対象語句の対訳と判断する第2の対訳判断手段

として機能させるプログラム。

[図1]



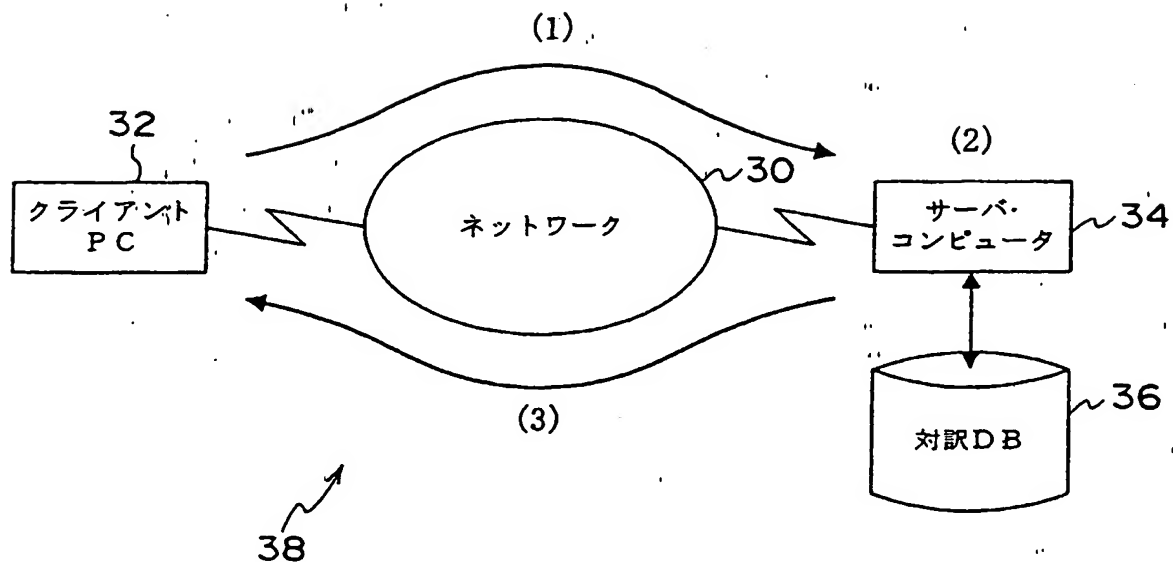
[図2]



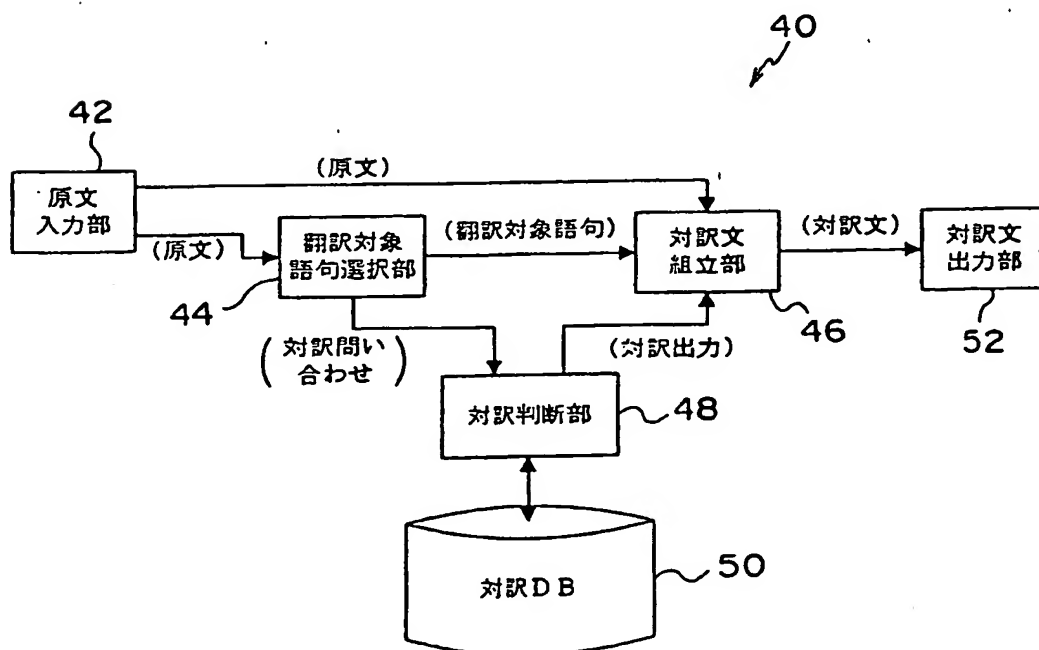
[図3]

	自然文	対訳文
1	operations management	業務管理
2	the operations of the new low	新法の適用
3	The system came into operation.	システムが稼働し始めた。
4	operations research	オペレーションズリサーチ
5	put into operation	実施する
6	After John's operation, his cancer was cured.	ジョンさんの手術後、彼の癌は治った。
7	operations room	作戦室
8	They monitor the computer operations in the cancer ward.	癌病棟のコンピュータ操作を監視する。

[図4]



[図5]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2004/015263

A. CLASSIFICATION OF SUBJECT MATTER
Int.Cl⁷ G06F17/28

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Int.Cl⁷ G06F17/28

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho	1922-1996	Toroku Jitsuyo Shinan Koho	1994-2005
Kokai Jitsuyo Shinan Koho	1971-2005	Jitsuyo Shinan Toroku Koho	1996-2005

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

JICST FILE (JOIS), WPI, INSPEC (DIALOG)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	JP 2003-263434 A (Kabushiki Kaisha Kokusai Denki Tsushin Kiso Gijutsu Kenkyusho),	1-2, 12-13,
Y	19 September, 2003 (19.09.03),	15, 16, 18
A	Par. Nos. [0021] to [0022] (Family: none)	3-10
Y	JP 04-160473 A (Hitachi, Ltd.), 03 June, 1992 (03.06.92), Page 4, lower left column, line 4 to page 4, lower right column, line 6 (Family: none)	11, 14, 17, 19
Y	JP 2000-020524 A (Fuji Xerox Co., Ltd.), 21 January, 2000 (21.01.00), Par. No. [0008] (Family: none)	3-10
		6

☐ Further documents are listed in the continuation of Box C.☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search
17 January, 2005 (17.01.05)Date of mailing of the international search report
01 February, 2005 (01.02.05)Name and mailing address of the ISA/
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

A. 発明の属する分野の分類 (国際特許分類 (IPC))

Int. Cl⁷ G06F17/28

B. 調査を行った分野

調査を行った最小限資料 (国際特許分類 (IPC))

Int. Cl⁷ G06F17/28

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	1922-1996年
日本国公開実用新案公報	1971-2005年
日本国登録実用新案公報	1994-2005年
日本国実用新案登録公報	1996-2005年

国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)

JICSTファイル (JOIS), WPI, INSPEC (DIALOG)

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
X	JP 2003-263434 A(株式会社国際電気通信基礎技術研究所) "2003.09.19, 【0021】 - 【0022】 (ファミリーなし)"	1-2, 12-13, 15, 16, 18
Y		3-10
A		11, 14, 17, 19
Y	JP 04-160473 A(株式会社日立製作所) 1992.06.03, 第4頁左下欄第4行-第4頁右下欄第6行 (ファミリーなし)	3-10
Y	JP 2000-020524 A(富士ゼロックス株式会社) 2000.01.21, 【0008】 (ファミリーなし)	6

☐ C欄の続きにも文献が列挙されている。☐ パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー

「A」 特に関連のある文献ではなく、一般的技術水準を示すもの

「E」 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの

「L」 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)

「O」 口頭による開示、使用、展示等に言及する文献

「P」 国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献

「T」 国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの

「X」 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの

「Y」 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの

「&」 同一パテントファミリー文献

国際調査を完了した日

17.01.2005

国際調査報告の発送日

01.2.2005

国際調査機関の名称及びあて先

日本国特許庁 (ISA/JP)

郵便番号 100-8915

東京都千代田区霞が関三丁目4番3号

特許庁審査官 (権限のある職員)

和田 財太

5M

9459

電話番号 03-3581-1101 内線 3597

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

☐ BLACK BORDERS

☒ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES

☒ FADED TEXT OR DRAWING

☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING

☐ SKEWED/SLANTED IMAGES

☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS

☐ GRAY SCALE DOCUMENTS

☒ LINES OR MARKS ON ORIGINAL DOCUMENT

☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY

☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.